Guidebook for Genetic Resources Documentation

A self-teaching approach to the understanding, analysis and development of genetic resources documentation

K.A. Painting, M.C. Perry, R.A. Denning and W.G. Ayad

International Plant Genetic Resources Institute
IPGRI

Guidebook for Genetic Resources Documentation

Guidebook for Genetic Resources Documentation

A self-teaching approach to the understanding, analysis and development of genetic resources documentation

K.A. Painting, M.C. Perry, R.A. Denning and W.G. Ayad

In consultation with:

Genebank Section, Field Crops Research Institute,
Bahteem Agricultural Research Station, Egypt

Universidad de San Carlos de Guatemala
Instituto de Ciencia y Tecnología Agrícola, Guatemala

The National Genebank of Kenya, Kenya

International Plant Genetic Resources Institute
IPGRI

The International Plant Genetic Resources Institute (IPGRI) is an autonomous international scientific organization operating under the aegis of the Consultative Group on International Agricultural Research (CGIAR). IPGRI's mandate is to advance the conservation and use of plant genetic resources for the benefit of present and future generations. IPGRI works in partnership with other organizations, undertaking research, training and the provision of scientific and technical advice and information, and has a particularly strong programme link with the Food and Agriculture Organization of the United Nations. Financial support for the agreed research agenda of IPGRI is provided by the Governments of Australia, Austria, Belgium, Canada, China, Denmark, France, Germany, India, Italy, Japan, the Republic of Korea, the Netherlands, Norway, Spain, Sweden, Switzerland, the UK and the USA, and by the Asian Development Bank, IDRC, UNDP and the World Bank. This work was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada.

Citation:
Painting, K.A., M.C. Perry, R.A. Denning and W.G. Ayad. 1995. Guidebook for genetic resources documentation. International Plant Genetic Resources Institute, Rome.

ISBN 92-9043-215-2
Reissued October 1995

IPGRI
Via delle Sette Chiese 142
00145 Rome
Italy

Contents

Chapter 4
Analysis of data generation and use in genebanks ▶ 57

Chapter 5
Data recording ▶ 91

Chapter ***10***
Implementation and maintenance of the system ▶ *219*

Figures and tables in this guidebook

Chapter 2: Introduction to genebanks and documentation systems ▶ 7

Chapter 3 Information processing in genebanks ▶ 37

Chapter 4
Analysis of data generation and use in genebanks ▶57

Chapter 5
Data recording ▶ 91

Chapter 6
Organisation of different types of data ▶ 109

Chapter 7
Computer basics ▶ 131

Chapter 8
Database basics ▶ 153

Chapter 9
Building the system ▶ 181

Preface

This *Guidebook for Genetic Resources Documentation,* the IBPGR Genebank Management System (GMS) software and the *GMS User's Guide* are three outputs of a project funded jointly by IBPGR and the International Development Research Centre (IDRC), Ottawa, Canada entitled 'Development of sustainable national agricultural genetic resources documentation programs: A self-teaching approach'. The primary focus of this project was to provide mechanisms that will promote self-sustainability of genetic resources documentation efforts by imparting and disseminating knowledge that can be used to develop the technical skills and thought processes necessary for the efficient documentation of genetic resources. The intention of IBPGR and IDRC has been to involve all willing participants in this project from its start. Three national plant genetic resources programmes were formally asked to consult with IBPGR on this project. In addition, during the early stages of the project, many other organisations collaborated with IBPGR by giving advice and holding discussions on the focus of the *Guidebook,* documentation methodology, and software design. Three regional workshops were held to encourage collaboration in genetic resources documentation and to assess this *Guidebook* and the associated software.

This *Guidebook* will assist users in making their own decisions on the analysis, design, implementation and use of both manual and microcomputer-based documentation systems. The GMS software and the *User's Guide* are provided with the *Guidebook* as an optional mechanism for the establishment of a documentation system.

This document represents a timely, self-teaching treatment of a subject of concern to those involved with many aspects of genetic resources documentation, characterisation, conservation, management and utilisation. As an essential part of the development of the project, the *Guidebook's* authors solicited the expert advice of reviewers and collaborators and responded to their valuable suggestions in the preparation of the final work. IBPGR is thankful for the range and extent of comments that many individuals from many organisations made on all the drafts of the *Guidebook.*

Acknowledgements

Project staff:

Mark Perry, project leader
Yawooz Adham
R.K. Arora
Franck Attere
George Ayad
Jan Engels
Henry Kamau
Luis López
Zhou Ming-De
Armando Okada
Kevin Painting, principal consultant
Froylan Rincón
Paul Stapleton
Jane Toll
Lyndsey Withers
Zhang Zongwen

Credits for the *Guidebook for Genetic Resources Documentation:*

Adriana Alercia (Spanish translation)
George Ayad (technical consultations and proof reading)
Margarita Baena (Spanish proof reading)
Isolde Canale (French translation and proof reading)
Rachel Denning (editing, writing, graphics, cover and book design)
Serena Lovell (typing and secretarial organisation)
Patti Martin (software design and production)
Ken Novak (software design and production)
Kevin Painting (principal writer, software outline and editing)
Mark Perry *(Guidebook* outline, project development, overall project management, editing and proof reading)
Froylan Rincón (Spanish translation)
Paul Stapleton (chief editor and production manager)
Lyndsey Withers (proof reading)

Consulting institutes:

Genebank Section, Field Crops Research Institute, Bahteem Agricultural Research Station, Bahteem, Egypt

Universidad de San Carlos de Guatemala, Instituto de Ciencia y Tecnología Agrícola, Guatemala City, Guatemala

The National Genebank of Kenya, Kikuyu, Kenya

Collaborating institutes:

Agicultural Research Institute, Nicosia, Cyprus

Centro Agronómico Tropical de Investigatión y Enseñanza, Turrialba, Costa Rica

Centro International de Agricultura Tropical/Southern Africa Development and Cooperation Council Bean Programme, Arusha, Tanzania

Centro International de Mejoramiento de Maíz y Trigo, Mexico City, Mexico

Coffee Research Station, Lyamungu, Tanzania

Horticultural Research Station, Tengeru, Tanzania

Instituto Nacional de Investigaciones Forestales Agropecuarias, Mexico City, Mexico

International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Kenya Agricultural Research Institute, Nairobi, Kenya

National Commission on Plant Genetic Resources, San José, Costa Rica

Tropical Pesticides Research Institute, Arusha, Tanzania

Universidad Autónoma Chapingo, Chapingo, Mexico

University of Costa Rica, San José, Costa Rica

Reviewers:

Y.J. Adham
E. Bettencourt
E. Bird
S. Bjarnason
S. Blixt
J. Chweya
J.M.M. Engels
L. Engle
M. Esquivel-Perez
E. Ferreira Rossi
B. Fraleigh
L. Guarino
J. Hanson
T. Hazekamp
V. Holubec
Z. Huaman
T. Hussein
M. Iwanaga
G. Juvik
H.N. Kamau
D. Kiambi
J. Konopka
L. López
D.W. Maina
W.Y. Marandu
P. Martin
D. Mbewe
Z. Ming-De
J. Mowder
X. Perrier
T.J. Ruredzo
E.N. Seme
J. Serwinski
N. Stavropolos
S. Steven
S. Taba
J.A. Toll
F. Vasquez
Z. Zhongwen

Workshop staff and participants

WEST ASIA AND NORTH AFRICA:

Workshop held at International Centre for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria (21-25 June 1992)

Workshop staff:

Y.J. Adham, IBPGR

J. Konopka, ICARDA

K.A. Painting, IBPGR

M.C. Perry, IBPGR

Participants:

L. Al Jarrah, Syria

A. Antypas, Syria

R. Attaie, Iran

M.S. Bhatti, Pakistan

A. Ghorbel, Tunisia

T. Hussein, Egypt

A. Karagöz, Turkey

G. Mir-Ali, Syria

S.B. Molatim, Lybia

LATIN AMERICA:

Workshop held at Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Mexico City, Mexico (27-31 July 1992)

Workshop staff:

S. Bojorges, IBPGR

K.A. Painting, IBPGR

M.C. Perry, IBPGR

F. Rincón, IBPGR

Participants:

J. Tilleria Aguilar, Argentina

C. Loaisiga Caballero, Nicaragua

A. Cubillos, Chile

J. Estrella, Ecuador

J. Sanchez Gonzalez, Mexico

E. Kornelius, Brazil

E. Tuk Mena, Costa Rica

M. Esquivel Pérez, Cuba

E. Martinez Tambito, Guatemala

AFRICA:

Workshop held at the National Genebank of Kenya (NGK), Kikuyu, Kenya (7-11 September 1992)

Workshop staff:

H.N. Kamau, IBPGR

D. Kiambi, IBPGR

K.A. Painting, IBPGR

E.N. Seme, NGK

L. Withers, IBPGR

Participants:

R. Akromah, Ghana

E. Dulloo, Mauritius

W .Y. F. Marandu, Tanzania

E. Ibrahim Mohamed, Sudan

G. Mwila, Zambia

G.N. Ngae, Kenya

N. ne Nsaka, Burundi

T.J. Ruredzo, Zambia

E. Sendek, Ethiopia

Chapter 1

Introduction: the purpose of this guidebook

Developing a new documentation system for a genebank or updating a system that already exists at your genebank can be a daunting thought. It is an enormous job and it can be difficult to identify where to start.

If you are reading this guidebook, it is likely that you intend to be involved in some way with developing a new documentation system for your genebank. You may already have experience of genebank documentation, in which case you may have some ideas about how to approach the design of a new system. But equally it could be an area that is completely new to you.

1

Who should use this guidebook

This guidebook is aimed at people with different backgrounds and experience who are faced with the common objective of developing a new, or improving an existing genebank documentation system.

You may be working at a small to medium sized genebank which has not previously developed a sophisticated documentation system, or a newly established genebank which has not yet clearly defined its activities. Equally, you might be working at a large genebank with an existing documentation system, but want to look at ways that the existing system can be improved to cope more efficiently with the large volumes of work.

If you have been given the task of designing a new system, you will probably have academic qualifications in a scientific or technical field. It will be useful if you have some experience of genebank activities, though this is not essential. This guidebook takes into account the fact that you may have no previous experience of this area, by discussing typical genebank activities in detail. Some experience of working with computers would also be an advantage, but again this is not essential as the guidebook contains a detailed chapter on computer basics.

2 Background: the problem

Genebanks will have problems managing their collections if they do not have up to date, accurate and reliable information stored in a systematic and easily accessible way. Such a documentation system allows genebanks to use information to plan their day to day activities and to maximise the use of their often limited resources. Also, without an effective documentation system, genebanks cannot develop with a clear sense of direction and they cannot easily communicate or collaborate with other institutions.

There are a number of factors which hinder the development of sustainable documentation systems in genebanks. These hindrances, some or all of which you might have encountered, are discussed below.

2.1 Lack of expertise in documentation

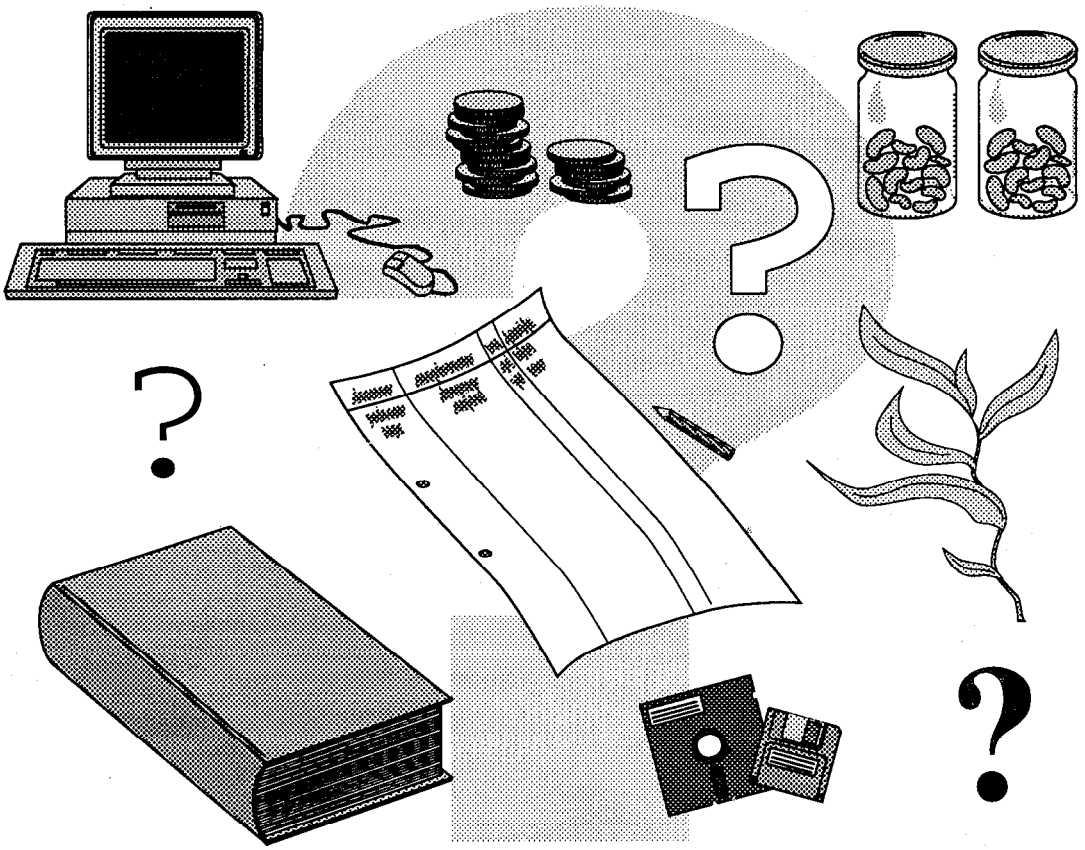
Documentation skills are an essential resource for any genebank. If there is a lack of expertise in documentation at your genebank, or if staff have not received adequate training, you will probably feel the need for advice and guidance on the major task of developing and implementing your documentation system.

2.2 Lack of training materials

If there are no available staff at your genebank with expertise in documentation to pass on their valuable knowledge, suitable training materials will be needed to guide you through the different stages of developing and implementing your new system. These self-teaching materials must be simple to follow, demanding little or no additional assistance. If you don't have access to such self-teaching materials, the process of developing and implementing a new documentation system will be problematic.

2.3 Shortage of staff

Often genebank staff have many other duties to perform and only a limited amount of time to devote to documentation. This is a common problem in many genebanks. If time spent on documentation is not allocated wisely, problems will be experienced in both the development and the operation of the documentation system.



2.4 Low priority given for documentation

When allocating priorities in any organisation, the importance of all activities taking place must be assessed before any decisions are made. Often, not enough time or resources are assigned to documentation in genebanks. This will inevitably lead to inadequate documentation systems being operated.

2.5 High turn-over of qualified personnel

Well qualified personnel, such as database specialists and programmers, may not stay in the same job for long periods of time. Often they will be only working on short-term contracts. This potential high turnover of documentation personnel can disrupt the continuity in the development of the documentation system. It can also cause operational problems resulting from a lack of knowledge base covering the system as a whole.

2.6 Lack of appropriate technology

If appropriate technology is not available, or unsuitable technology is used, a documentation system will be difficult to operate. Problems will also arise when the system is modified or developed to meet the changing documentation and information needs. This would be true if, for example, a manual system is implemented when a computerized system is needed, or when an inappropriate computerized system is developed. Decisions made concerning what technology is used are usually related in some way to what budget is available for a particular project. However, foresight is essential when making such decisions. The equipment you invest in now will be with you for several years to come and the cheapest option is often not the best.

All these factors can contribute to a lack of documentation, or the development of documentation systems which are either inappropriate or inadequate for the needs of a genebank. What can be done to overcome this?

3 The solution: self-teaching materials

Training in documentation skills and the appropriate technology is vital to the development of a sustainable documentation system. Since training in documentation skills through personal tuition is not always available, self-teaching materials provide an excellent alternative.

This self-teaching guidebook for plant genetic resources documentation specifically addresses these training needs. It is designed to help you set up a manual or computerized documentation system appropriate to your requirements and resources. This is achieved by guiding you through all the areas you need to consider, thus putting you in a position where you are able to assess your own needs. It approaches the three major stages of setting up a manual or computerized documentation system in a step by step fashion through:

- ***Analysis*** of genebank activities to determine information and documentation needs
- ***Design*** of the manual and/or computerized system based on documen tation and information needs
- ***Implementation*** of the system that has been developed

4 Structure of the guidebook

The chapters of this guidebook form a progressive series of self-contained modules. They approach the different stages of construction of your documentation system in a logical order, discussing the basic documentation activities and providing training in the use of computers.

Central to this guidebook are two detailed questionnaires which are designed to analyse your genebank documentation and information needs. You should spend some time completing these, as they form a valuable foundation for use when designing the structure of your new documentation system. The results of these questionnaires are built on as you progress through the guidebook.

Exercises are also included in most chapters to test the comprehension of new topics.

The focus of each chapter is as follows:

Chapter 2; Introduction to genebanks and documentation systems, introduces you to the organisation and operation of genebanks and genebank documentation systems and describes the stages in building a system.

Chapter 3; Information processing in genebanks, describes the central role played by genebank procedures in the generation and processing of information and highlights the use of accession numbers, batch references and scientific names in building the genebank documentation system.

Chapter 4; Analysis of data generation and use in genebanks, looks in greater detail at many of the commonly performed genebank procedures with regard to data generation and use.

Chapter 5; Data recording, looks in more detail at how data can be recorded and how this affects data handling and information retrieval.

Chapter 6; Organisation of different types of data, deals with the use of manual forms in both computerized and manual systems and looks more closely at the organisation and operation of a manual documentation system.

Chapter 7; Computer basics, introduces the basic features of computers – what they consist of, how they are operated and their use as tools for carrying out a wide range of activities.

Chapter 8; Database basics, looks at the ways you can organise computer files based on sets of descriptors identified in the genebank analysis.

Chapter 9; Building the system, looks at the different stages of building a computerized documentation system based on the analysis you have performed at your genebank.

Chapter 10; Implementation and maintenance of the system, discusses the implementation of your documentation system, considerations for data security, procedures for changing your system and system documentation.

5 How to use this guidebook

It is important that the chapters of this guidebook are studied in sequence. The earlier Chapters 2 to 5, concentrate on genebank procedures. You are strongly advised to work through these chapters, even if you already have experience in this area. You will no doubt progress more rapidly than those with little or no experience in genebank procedures. Above all, you should not omit to work through the questionnaires and analyses as these are central to the development of your new documentation system.

If after analysing the questionnaire completed in Chapter 2, you come to the conclusion that a manual system is the right system for your genebank, Chapters 7 through 9 can be omitted as they concern the implementation of a computerized system. If, however, you are in any doubt about what type of system is most suitable for your genebank you are strongly advised to work through these chapters. Some of the points discussed concerning the benefits of a computerized system might cause you to rethink your options. Chapter 7 is intended as an introduction for people with little or no computing experience.

Any questionnaires should be filled in and analysed and exercises in comprehension should be completed satisfactorily *before* moving on to the next chapter. The sections dealing specifically with the operation of computerized documentation systems assume that the concepts introduced in previous chapters are thoroughly understood and that a full analysis of genebank documentation and information needs has been performed.

Working thoroughly through this guidebook will provide you with many more benefits than just the knowledge needed to design and implement a new documentation system. The process will give you greater insight into all aspects of how your genebank operates. The knowledge and skills gained will be a valuable asset to both you and your genebank for any future work you undertake.

Chapter 2

Introduction to genebanks and documentation systems

Chapter 2 will introduce you to the organisation and operation of genebanks and genebank documentation systems. When you have finished this chapter, you will be able to:

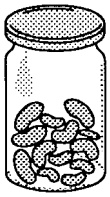
- Explain how genebanks differ from one another
- List four major categories of genebank
- Review different types of collection
- Distinguish data from information
- Describe the role information has in setting priorities
- Demonstrate the need for a documentation system
- List desirable features of a documentation system
- Describe how data are organised in a documentation system
- List the stages in the construction of a documentation system
- Obtain background information on your genebank using the questionnaire included

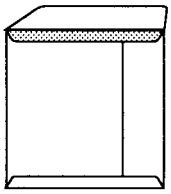
1 Introduction to genebanks

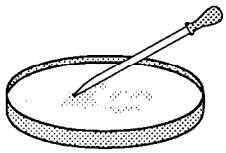
1.1 Genebanks as genetic resource centres

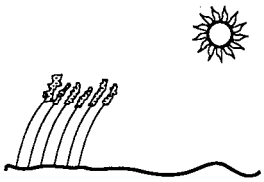
In the same way that libraries are resource centres for information, so plant genebanks are resource centres for living plant material. Plant genebanks maintain collections of plant material-with the aim of keeping them alive and preserving their characteristics for the future benefit of mankind and the environment. Genebanks are also called "plant genetic resources centres", emphasising that plants are sources of genetic characteristics – sources of diversity. The plants maintained include economically important food crops (modern and primitive cultivars and their weedy and wild relatives), horticultural plants, forages, medicinal plants and trees. Genebanks do not maintain all types of plants – this would be too great an undertaking. Instead, they are selective about the plants they keep.

The material that is actually maintained – whether it's seeds, tissue culture or actively growing plants – is called germplasm. Germplasm is the genetic material responsible for a plant's characteristics, transmitted from one generation to the next. It doesn't matter in what form the germplasm is maintained for the future so long as the aim of keeping the plant material alive and preserving their characteristics is achieved.









Although all genebanks conserve germplasm, they are much more than just cold rooms full of seeds. Genebanks are involved in a range of other activities which make the germplasm more useful to other scientists. Some of these activities are listed below.

- Acquisition of new germplasm samples
- Multiplication/regeneration of germplasm
- Characterisation and preliminary evaluation of germplasm
- Documentation and exchange of information on germplasm
- Conservation of germplasm
- Supply of germplasm
- Collaboration with other plant genetic resource centres
- Organisation of technical meetings and training workshops
- Research (e.g. germplasm enhancement, seed physiology)

We'll be looking in greater detail at some of these activities in later chapters.

1.2 Why do genebanks differ?

Genebanks differ from one another in their activities and how the activities are organised and performed.

Consider the differences between a genebank attached to a particular breeding programme and a national genebank. Their objectives, activities and-priorities are different. The breeding programme probably keeps germplasm for the few crops in the programme and since its major objective is use of the germplasm, it probably stores the germplasm under medium-term storage conditions. On the other hand, the national genebank generally keeps a wide range of species and, as one of its major objectives is *conservation,* it stores the germplasm under long-term

storage. The breeding programme also has different priorities with regard to the supply and receipt of germplasm; it probably wasn't equipped for distributing germplasm or having the necessary quarantine facilities to deal with batches of incoming seed. The reason the breeding programme wasn't equipped to do these things was because it doesn't need to do them.

1.3 Genebank aims and objectives

What a genebank is set up to achieve can be summarised in a short statement of its aim. Examples could be: "... to conserve barley germplasm for the in-house breeding programme"; "... to conserve national maize germplasm for the benefit of future generations and to ensure the supply of the germplasm without prejudice to all scientists." The genebank aim is often accompanied by a number of objectives. For example, a national genebank might have the following objectives:

- Long-term conservation of national plant genetic resources
- Germplasm regeneration
- Characterisation and evaluation work on specific germplasm
- Organisation of germplasm exploration and collection at a national level
- Introduction of germplasm
- National and international exchange of germplasm and information
- Training, education, and organisation of technical meetings and workshops

In order to achieve these objectives a number of different genebank activities must be performed. For example, the objective "Acquisition of wild *Arachis* germplasm" involves the activities illustrated in fig. 1.

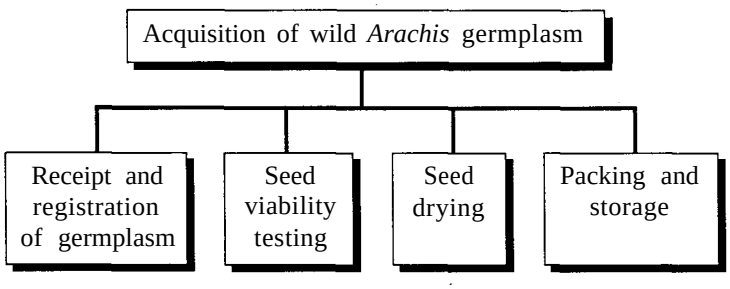


Fig. 1. Activities involved in the acquisition of wild *Arachis* germplasm

Many genebank activities generate and use data in their operation. It follows that any change in genebank objectives will have an effect on the activities that are performed and any data that are produced or used.

1.4 Different categories of genebanks

Four major categories of genebanks can be identified according to their aims.

Definitions:

INSTITUTIONAL GENEBANK

An institutional genebank is set up to conserve only the germplasm which is used in (or is potentially useful to) the research programmes in its host institute or host agricultural research centre.

NATIONAL GENEBANK

A national genebank is set up as a national plant genetic resources centre maintaining many different germplasm samples of current and potential interest for people working in plant research nationally. Commonly it will contain germplasm which has been collected nationally. It may be closely associated with a research programme or undertake its own research.

A national genebank can be a collaborative venture between national institutes or under the responsibility of one institute which collaborates with other national institutes.

REGIONAL GENEBANK

A regional genebank is set up as a collaborative venture between a number of countries in the same geographical region to conserve the germplasm from that region and to support plant research. Examples of regional genebanks include the Southern African Development Coordination Conference (SADC) Regional Genebank in Zambia, the Centro Agronómico Tropical de Investigatión y Enseñanza (CATIE) in Costa Rica and the Nordic Genebank for Agricultural and Horticultural Plants (NGB) in Sweden.

INTERNATIONAL CENTRES

Most of the international agricultural research centres (IARCs) of the Consultative Group on International Agricultural Research (CGIAR) have substantial germplasm collections, centered on particular crops called mandate crops, but also holding other crops (see table 1). Much of the germplasm is collected worldwide with international collaboration and is conserved for the benefit of plant genetic resources activities worldwide.

Genebanks don't usually work in isolation from other genebanks or genetic resources programmes. Often there will be collaboration between institutional and national genebanks as part of a national programme of plant genetic resources work. There will also be good contacts between the national genebanks and the international centres. This sort of cooperation between genebanks stimulates mutual development and ensures an effective international operation for the conservation of plant genetic resources.

Table 1. International agricultural research centres of the CGIAR and their mandate crops

INTERNATIONAL CENTRE	GERMPLASM HOLDINGS
Centro Intemacional de Agricultura Tropical (CIAT), Colombia	*Phaseolus vulgaris*, *Phaseolus lunatus*, other *Phaseolus* species, cassava, wild cassava *(in vitro)*, forage grasses, forage legumes
Centro Intemacional de Mejoramiento de Maíz y Trigo (CIMMYT), Mexico	wheat, maize, triticale, rye, barley
Centro Internacional de la Papa (CIP), Peru	potato, wild species of potato, sweet potato, wild species of sweet potato, other Andean root and tuber
International Centre for Agricultural Research in the Dry Areas (ICARDA), Syria	cereals, food legumes, forages
International Centre for Research in Agroforestry (ICRAF), Kenya	agroforestry species
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India	*Sorghum*, finger millet, pigeonpea, foxtail millet, little millet, barnyard millet, pearl millet, chickpea, groundnut, proso millet, sawa millet, kodo millet
International Institute of Tropical Agriculture (IITA), Nigeria	sweet potato, cassava, yams, *Musa* spp., taro, wild *Vigna*, multipurpose trees, bambara groundnut, misc. food legumes, cowpea, soybean, rice
International Livestock Centre for Africa (ILCA), Ethiopia	grasses, legumes, browse
International Network for the Improvement of Banana and Plantain (INIBAP), France	banana, plantain
International Rice Research Institute (IRRI), Philippines	*Oryza sativa* (Asian rice), *O. glaberrima* (African rice), wild species and species hybrids, taxa in general related to *Oryza*
West Africa Rice Development Association (WARDA), Côte d'Ivoire	rice

1.5 Different types of collections in genebanks

We've been talking about the different types of genebanks but not much about how the germplasm is maintained. Germplasm is maintained in a number of different "collections"; these collections are used by the genebank in different ways. There are three types of collections that you'll commonly come across in the literature: base, active, working. The purpose for setting up each collection type is quite different. Some useful working definitions are now given.

Definitions:

BASE COLLECTION

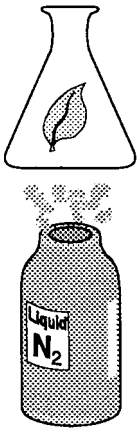
A collection of germplasm which is kept for long-term, secure conservation and is not to be used as a routine distribution source. Seed is usually stored at sub-zero temperatures with low seed moisture content.

ACTIVE COLLECTION

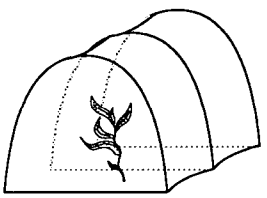
A collection of germplasm used for regeneration, multiplication, distribution, characterisation and evaluation. Ideally germplasm in the active collection should be maintained in sufficient quantity to be available on request. Active collection germplasm is commonly duplicated in a base collection and is often stored under medium to long-term storage.

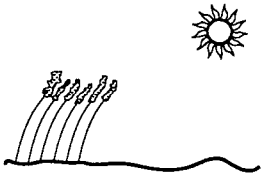
WORKING COLLECTION

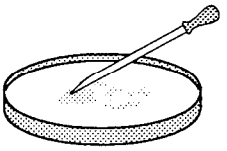
A collection of germplasm which is used by breeders or researchers in their work. Conservation is not a priority in working collections. An example of a working collection might be germplasm samples derived from an active collection being used by breeders in the course of their breeding programme.



You may also come across the terms "field collection" and "*in vitro* collection". A field collection (or field genebank) is a collection of living plants (e.g. fruit trees, glass house crops and field crops). Germplasm which would otherwise be difficult to maintain as seed can be kept in field collections. Often it is the only type of collection maintained by the genebank. Field collections should be thought of as active collections. An *in vitro* collection is a collection of germplasm kept as plant tissue grown in active culture on solid or in liquid medium. Like field collections, these should be thought of as active collections. In some cases the tissue is stored at very low temperatures such as in liquid nitrogen. These collections can be considered as base collections.







You should not confuse the type of collection with the storage conditions used. Storage conditions are defined more specifically as:

- Long-term storage: -1 to -20°C, 4-6% seed moisture content, more than 10 years storage
- Medium-term storage: 1-10°C, 15% seed moisture content, up to 10 years storage
- Short-term storage: reduced seed moisture content, ambient temperature

Although many base collections are maintained under long-term storage conditions, you might find them maintained under medium-term or even short-term storage conditions particularly where facilities such as cold rooms or deep freezers are not available. (Some people might argue that this does not constitute a base collection). You'll also find active collections maintained under long-term storage conditions.

Genebanks commonly maintain their germplasm in more than one collection or under more than one storage condition. An unpublished survey by IBPGR of 321 genebanks in 1991 showed that 103 of these maintained germplasm under more than one storage condition (long, medium, short-term storage). Genebanks choose to use more than one method of maintaining their valuable germplasm principally as a safeguard against loss but also to simplify working practices.

1.6 Genebank organisation

The larger genebanks are generally staffed by a curator (genebank manager) and scientists working in the main activities together with support staff. This organisation is illustrated in fig. 2.

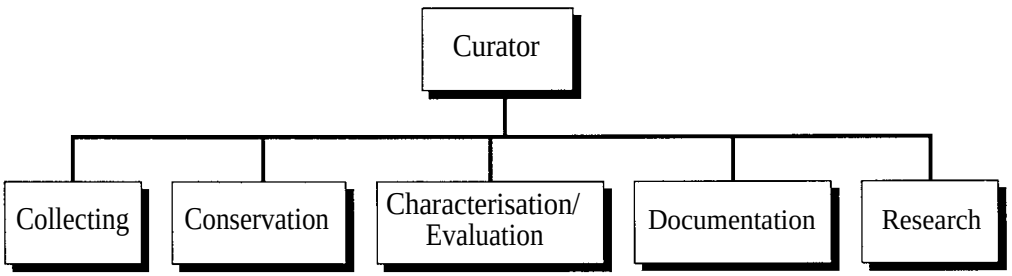


Fig. 2. Organisation of activities in a genebank

The different units have specific objectives for their work which are determined by the overall objectives of the genebank. Usually, the documentation unit works closely with the other units supporting their activities.

2 Introduction to documentation systems

2.1 Data and information

Definitions:

DATA

Quantitative or qualitative values derived from observations.

INFORMATION

The meaning that emerges when data are recorded, classified, organised, related or interpreted.

DESCRIPTOR

An identifiable and measurable characteristic (e.g. silique attitude, see fig. 3) used to facilitate data classification, storage, retrieval and use.

DESCRIPTOR LIST

A collation of all the individual descriptors used for a particular species or crop.

DESCRIPTOR STATE

A clearly definable state which a descriptor can take (e.g. 3 Erect, 5 Hanging or 7 Pointing down, see fig. 3).

DOCUMENTATION SYSTEM

Any way of storing and maintaining data. A documentation system can use manual methods (such as hand written records) and/or completely computerized methods for data storage and maintenance. The system is also designed for information retrieval.

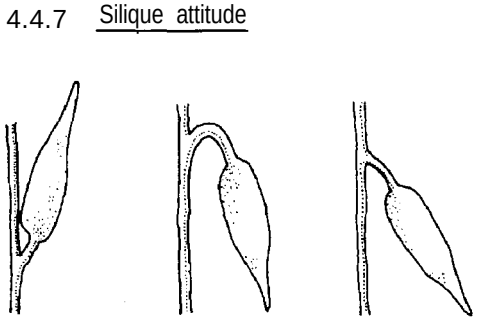


Fig. 3. Different descriptor states for silique attitude in *Brassica*

When you weigh a batch of seeds or measure the height of a plant or describe the colour of flowers, you are making observations on the plant characteristics. In plant genetic resources work, it is customary to refer to these characteristics as *descriptors*. What you actually write in your laboratory notebook when making these observations is data. Data can be *quantitative* and dealing with numbers (e.g. 94g, 34mm, 67%) or they can be *qualitative*, that is, a description of the object being examined (e.g. brown, hairy, horizontal). For many descriptors there are a limited number of *descriptor states* (see table 2). For instance the descriptor "plant growth habit" might only have three descriptor states: procumbent, decumbent and erect. In other cases, the descriptor states are continuously variable such as seed weight or plant height. Some further examples are given in table 2.

Table 2. Examples of different descriptors and their descriptor states

DESCRIPTOR	DESCRIPTOR STATES
Flower colour	white, cream, yellow, orange, green, dark green, red, dark red
Collection source	wild habitat, farm land, farm store, backyard, village market, commercial market, institute, other
Altitude of collecting site	continuously variable
Monthly rainfall	continuously variable
Seed germinating in ripe fruit	absent, present

Information is not the same as data although people sometimes use the terms interchangeably. Information has meaning whereas data do not. For example, what does "94g" mean – 94g of what? Table 3 highlights this essential difference.

Table 3. The difference between data and information

DATA	INFORMATION
94g	The weight of a seed batch in the cold store is 94g
34mm	The rainfall in June 1991 was 34mm
1324	Accession number 1324
67%	The viability of a seed batch is 67%
brown	The colour of seeds in a seed batch is brown

From table 3 you can see that when data are manipulated in some way, described, or compared with other data, some sort of meaning emerges. This is the essence of information – it has meaning. Data can be recorded in a number of different ways, but the way you actually record the data can profoundly affect the usefulness of the information you retrieve. This is an important point and we will be returning to it in Chapter 5.

Remember that the information is only as good as the data – if the balance used in the genebank aren't working properly, can you really trust information on seed weight?

2.2 The use of information for genebank management

Much of the information acquired and generated by a genebank is of interest and of value to the scientific community. Genebanks will commonly distribute their accessions together with relevant information such as passport information and information obtained from characterisation and preliminary evaluation trials. Information from specific studies will often be published in reports, catalogues or journals.

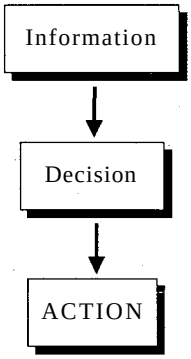
Not all information produced in genebank activities is of interest to other scientists but it is of vital importance in genebank management for:

- Setting priorities
- Planning activities
- Managing resources

Without a supply of up to date, accurate and reliable information, a genebank cannot carry out *any* of its activities effectively – not even its vital conservation work. Why is this?

A genebank only has limited resources (staff, equipment, experimental plots) and limited time. The curator therefore has to set priorities and decide which activities are more important than others at a particular time. Problems will arise if the wrong priorities or unrealistic objectives are set. So, how does the curator set the right priorities? How does the curator decide?

In all decision making, *information* is needed on which to base the decision. Moreover, the information must be accurate, reliable and up to date otherwise the wrong decision could be made.



Consider germplasm regeneration. Not all accessions can be regenerated each season so the curator has to set priorities. Questions which might be asked include:

- Which samples must, at all costs, be regenerated?
- Which regenerations are less urgent?
- What are the consequences of not regenerating certain samples?

To help answer these questions, the curator needs information about seed stock levels, seed viabilities and how frequently particular accessions are distributed. Only when this information is obtained can a wise decision be made about regeneration priorities. Next, further informa-

tion is needed to plan and organise the work. The curator would need information on:

- How many accessions are to be regenerated
- How much land is required
- Where the work will be carried out
- How many accessions will be regenerated at each site
- How many staff members are required to do the work
- Which descriptors will be measured/observed

With this information the work schedule can be planned.

It's clear that information plays an important role in the planning and organisation of genebank activities. What would happen if the information were not accurate, or if it were out of date? The wrong decisions could be made. If the wrong decisions were made when planning regenerations, valuable germplasm might be lost.

2.3 The need for a documentation system

A genebank needs a constant supply of accurate, reliable and up to date information to function efficiently. Where can it get this information? The genebank would not use the human memory as the only source of information as it's not always completely reliable - people can easily forget facts or confuse details. In practice, the genebank will rely on a documentation system as a source of information to assist in the planning and operation of many genebank activities.

However, a documentation system isn't only used for information retrieval. It is also used for:

- Data storage
- Data maintenance (updating existing data)
- Data processing and analysis
- Data exchange

Genebanks differ from one another according to their activities and how the activities are organised. Since documentation systems support all these activities, it follows that the documentation systems operated by separate genebanks will also be different. Many genebank documentation systems do show *some* similarity in design and operation, but each will be different as they are tailor-made according to the documentation and information needs of the genebank.

2.4 Desirable features of a documentation system

2.4.1 Data integrity

Information retrieved from a documentation system must be accurate, reliable and up to date for it to be of value – otherwise what's the point of using the documentation system at all? You'll have to think carefully how the system can be designed and operated to facilitate maintenance of accurate information.

2.4.2 Fast information retrieval

If your system is well designed, retrieving information will be a simple and straightforward process. If it isn't well designed you may spend hours, or worse still, not be able to supply the information at all! If similar information is needed on a regular basis and it takes you several hours to locate the information each time, you're spending far too much of your time on information retrieval. Remember – the documentation system is working for you, not the other way round!

2.4.3 User-friendly operation

Data do not appear in a documentation system as if by magic - the chances are that you or a colleague will have entered the data. In fact you'll be spending a fair amount of your time at the computer or writing up forms so anything that reduces this workload is really helpful. It helps enormously if the documentation system is user-friendly. A user-friendly system can be used with the minimum of training. One of the commonest problems with documentation systems is that they are difficult and cumbersome to use. Remember – the documentation system is working for you, not the other way round. If the documentation system is user-friendly, you'll find considerably fewer errors creeping in and the system will be a lot more popular with other people.

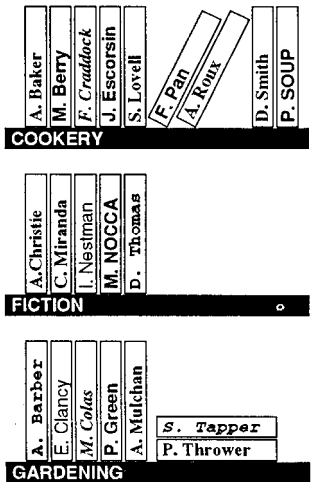
2.4.4 Flexible operation

The documentation system should not be rigid in its operation. It should be able to cope with different requests for information and accommodate changes in genebank procedures. If a new curator was appointed, would the curator have the same information needs? Probably not. If there were a change in genebank objectives or a new procedure, would this affect the documentation system? Yes it would! Therefore, you should try to anticipate information requests and changes in genebank procedures as far as possible.

2.4.5 Organisation of data

Data are not stored in a documentation system in a random, chaotic way; if they were, maintenance of the system would be tedious and information retrieval would be impossible. Instead, the data are organised into groups which are *practical to use* – practical for data recording, storage and maintenance and practical for information retrieval. In a genebank, practical groups are closely associated with genebank procedures e.g. seed viability testing, characterisation of wild *Arachis,* and seed moisture content determination.

To Clarify



Next time you go into a book shop, look at how the books are arranged. It's unlikely that they will have been arranged in strict alphabetical order of the author's name. It would be confusing (and irritating) to find cookery books, detective novels, atlases, biology text books all mixed together. Usually they will be arranged according subject area (e.g. fiction, non-fiction, reference) and then further organised into alphabetical order. If it's a large book shop, they have probably sub-divided the major groups – fiction divided into crime, romance, action, science fiction, childrens' books and so on. The groups into which the books have be organised are useful to the person buying books - the person can easily find a book on a particular subject. If the bookseller decided instead to arrange the books according to the colour of the book covers, the books might look attractive on the shelves but it wouldn't be of much use to-the shopper!

Using this example, you can see how the different items (in this case, books) can be arranged into practical groups. These groups are practical because some thought has been given to different peoples' needs; in this case, the need is for the subject of the book.

In a genebank documentation system, users needs must be taken into account when organising data into practical groups. The groups defined must be practical for data recording, storage and maintenance and practical for information retrieval. Sometimes a group designed for information retrieval is not always practical for data-recording, storage and maintenance; this is particularly true of manual systems. Part of the skill in designing a documentation system is to define groups which are practical for all operations that are likely to be performed. Luckily, these practical groups are usually associated with the genebank procedures themselves. We will be looking more closely at genebank procedures and how you can define these groups in the next two chapters.

2.5 Stages in construction of a documentation system

The construction of a documentation system requires detailed analysis and planning before the design of any manual forms and/or databases. In fact the process should be heavily slanted towards detailed analysis if a flexible and user-friendly system is to be implemented.

Six stages can be identified as follows.

2.5.1 Stage 1: Obtaining background information about the genebank

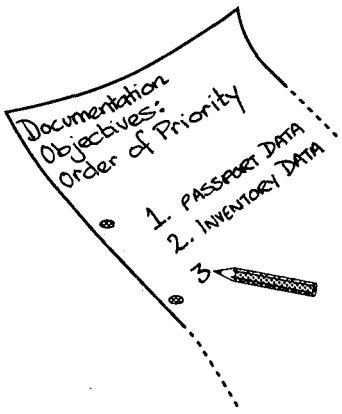
Purpose: to gather essential background information on the set-up of the genebank which will help define documentation objectives and facilitate resource management.

In order to develop a documentation system tailored to the genebank's needs, you need to analyse the genebank's set-up, and resources with the full cooperation of other staff members. This will give you essential information about the genebank from which you can later develop documentation objectives. It will also help you make decisions on how best to use available resources.

This analysis is covered in section 5 of this chapter.

2.5.2 Stage 2: Define documentation objectives

Purpose: to define priority areas for the documentation effort.



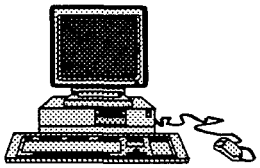
You need to have a clear idea of the areas of genebank work that need documenting and the priorities for documentation. You should write these down as documentation objectives and list them in order of importance. These objectives could include documentation of passport data, inventory data, seed handling procedures, distribution data, characterisation and evaluation trials. The objectives could also include dissemination of information.

Setting documentation objectives is easy if you are told by the genebank curator which areas to cover and which areas are more important than others. You'll probably set more documentation objectives than you can cope with at the same time – that's why you have to make priorities. It's important that you identify the priority areas right from the start to avoid problems later on.

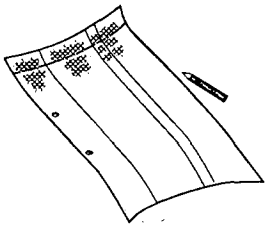
Also, remember that it's important to identify whether documentation of certain data is essential. Your time is precious so you don't want to spend time documenting unimportant data unless, of course, you like creating unnecessary work!

2.5.3 Stage 3: Analysis of genebank procedures

Purpose: to identify documentation requirements of each procedure and the relations between different procedures



VS



Having defined your documentation objectives you can start the detailed analysis of relevant genebank procedures. This should explore the resource requirements of each procedure, and the different types of data which are generated or used in each procedure. This will help you later on in deciding how best to handle the data e.g. the suitability of computers and/or manual forms.

The analysis will show how the procedures relate to one another. This information can be used to build up a flow chart showing the relationship between genebank procedures and information flow. This will help you later on in deciding how best to handle the data and will help you define documentation procedures.

The analysis will also give you an idea of the information requirements of the system. It's useful if you can build up a list of questions that are asked of the documentation system by the users. These questions will help you later on in deciding how best to arrange the data in the documentation system for flexible and effective information retrieval and help you define documentation procedures.

This analysis is described in Chapters 3 and 4.

2.5.4 Stage 4: Identify meaningful sets of descriptors

Purpose: to identify meaningful sets of descriptors from the analysis of procedures.

Much of the data you'll be recording will be concerned with individual genebank accessions. To facilitate the operation and maintenance of the documentation system, you need to organise the descriptors into practical sets. You can think of these sets as separate books, folders or forms in a manual system or separate files in a computer system (e.g. "characterisation of wild *Arachis*", "viability testing"). These sets are practical in terms of how data are recorded and used and practical in terms of information retrieval. This is covered in more detail in Chapters 4 and 5.

2.5.5 Stage 5: Develop data formats and manual forms and/or computer screen entry forms

Purpose: to develop formats for data which facilitate data recording and flexible information retrieval and to develop manual forms and/or computer database formats and screen entry forms that will be used at each stage of the documentation process.

An important task of the documentation specialist is the design of manual forms and computer screen forms for straightforward data entry which minimise, the risk of errors creeping into the system. If these forms are designed well, the accuracy of data entry is ensured. These design considerations are covered in Chapter 6 (for manual systems) and Chapters 8 and 9 (for computerized systems).

2.5.6 Stage 6: Develop documentation procedures and implement the new system

Purpose: to develop documentation procedures to facilitate operation of the system and to implement the system.

A documentation system can be well thought out and well constructed but not used unless there are clearly defined documentation procedures and training is given in the use of the system. These considerations are discussed in Chapter 6 (for manual systems), Chapters 8 and 9 (for computerized systems) and Chapter 10 (general considerations).

2.6 Where to go from here

You should now attempt the exercises in the following section which test your comprehension of the concepts introduced in this chapter. When you have successfully completed these exercises you should proceed to the questionnaire. This questionnaire is designed to gather essential background information about your genebank which we'll use in the design of the documentation system.

3 EXERCISES

Exercises

Answer as many of the following questions as you are able. Refer back to the text if you need assistance.

1. Indicate whether the following statements are true or false:
   a. Germplasm is maintained in genebanks
   b. All genebanks maintain seeds
   c. All genebanks perform the same activities
   d. Some genebanks maintain more than one collection
   e. All genebanks maintain germplasm under long-term storage conditions
   f. Base collections are not used for distributing seed
   g. Active collections contain actively growing plants
   h. Field collections maintain field crops
   i. The aim of conservation is to preserve the germplasm and its characteristics
2. Explain why genebanks are set up and what they maintain.
3. Explain why genebanks differ from one another.
4. Review the different types of collections in genebanks with specific reference to the storage conditions used.
5. Indicate whether the following statements are true or false:
   a. Data are qualitative or quantitative values
   b. Information on its own is meaningless
   c. All documentation systems use computerized methods
   d. Although not all documentation systems in genebanks are the same, they should be
   e. Information is used in making decisions
   f. Documentation systems are only used for getting information
   g. Documentation systems should be flexible in operation
   h. Data in a documentation system should be user-friendly
   i. Data in a documentation system can be organised into groups
6. What is the difference between data and information? Give examples of each.
7. Why is it necessary to set priorities in genebank work?
8. Explain the role information plays in managing genebank activities.
9. Explain why documentation systems differ between genebanks.
10. Describe some desirable features of documentation systems. Explain why these features are desirable.

EXERCISES

11. Discuss how data are organised in a documentation system and how this facilitates the use of the system.

12. Describe the different stages in construction of a documentation system. What is the purpose of each stage?

4 Questionnaire: Obtaining background information on your genebank

4.1 Purpose of the questionnaire

The purpose of the following questionnaire is to help you find out essential background information about your genebank before you start the detailed analysis of your genebank procedures in Chapter 4. Information gathered here is of a general nature and will highlight areas for further study in the detailed analysis. These include:

- The activities that the documentation system might deal with
- The need for operating a documentation system on more than one site
- The need for data exchange
- The extent to which standard descriptor lists need to be used
- Areas which need particular care with regard to the design of documentation procedures

If you work in plant genetic resources but do not work in a genebank, you will still find it useful to go through this questionnaire as many of the questions are relevant to your situation.

4.2 How to fill in the questionnaire

Set time aside to work thoroughly through this questionnaire and analysis. The results will give essential input and recommendations for the design of your documentation system.

As a general rule, it's important that staff members are consulted at every stage of the analysis and design process. After all, these people will probably be using the documentation system for recording data or for information retrieval in the future. If they are consulted, they will feel more directly involved and happier with the final documentation system when it is eventually introduced.

You will find it useful to consult the curator when filling in this questionnaire, particularly if you are fairly new to the genebank or you do not know the answers to some of the questions.

The questionnaire is divided into three sections, each gathering information in the following areas:

1. Your genebank's relation with other genetic resources programmes
2. Your genebank's aim and areas of activities
3. Your genebank's organisation

You should photocopy this questionnaire so that you have a separate sheet to refer to as you work through the guidebook.

At this stage, you should not write anything in the right hand column headed "Analysis". Use only the central area for your responses. You will find an example of a completed questionnaire in appendix 1.

QUESTIONNAIRE: YOUR GENEBANK

NAME OF GENEBANK:

LOCATION OF GENEBANK:

YOUR NAME:

DATE:

SECTION I: YOUR GENEBANK'S RELATION WITH OTHER GENETIC RESOURCES PROGRAMMES

1. *Which category does your genebank fit into?* **ANALYSIS**

- Institutional ☐
- National ☐
- Regional ☐
- International ☐
- Other (please specify) ☐ _____

2. *When was your genebank established?* **ANALYSIS**

- In the planning stage ☐
- Less than a year ago ☐
- 1-2 years ago ☐
- 3-5 years ago ☐
- More than 5 years ago ☐

3. *Is there a national programme of plant genetic resources activities in your country?* **ANALYSIS**

- Yes or in the planning stage ☐
- No ☐

4. *Does your genebank collaborate with other genetic resource programmes in other institutes or genebanks?* **ANALYSIS**

- Yes ☐
- No ☐

SECTION II: YOUR GENEBANK'S AIM AND AREAS OF ACTIVITIES

5. *Summarise in a short phrase your genebank's aim* ANALYSIS

[In this phrase try to explain what the genebank has been set up to achieve]

6. *Which of the following activities does your genebank undertake? (tick as many as appropriate)* ANALYSIS

- ☐ Germplasm collecting
- ☐ Acquisition of new germplasm samples
- ☐ Multiplication/regeneration of germplasm
- ☐ Characterisation and evaluation of germplasm
- ☐ Maintenance of germplasm
- ☐ Germplasm health/quarantine
- ☐ Dissemination of information on germplasm
- ☐ Selection of germplasm for distribution
- ☐ Research (e.g. germplasm enhancement, seed physiology)
- ☐ Organisation of technical meetings/training workshops

7. *In which form(s) is/are the germplasm maintained and in which type of collection?* ANALYSIS

	Base	Active	Working
Seeds	☐	☐	☐
Field genebank/collection	☐	☐	☐
In vitro	☐	☐	☐
Pollen	☐	☐	☐

8. *How many accessions are maintained by your genebank?* ANALYSIS

- ☐ Less than 100
- ☐ 100-500
- ☐ 500-1,000
- ☐ More than 1,000

9. *How many different species are held?* ANALYSIS

- ☐ 1
- ☐ 2-10
- ☐ More than 10

10. *Is your genebank in an expanding phase?* **A** *N A L Y S I S*

Yes ☐
No ☐

11. *Approximately how many samples do you distribute in a typical year?* **A** *N A L Y S I S*

Fewer than 100 ☐
100-500 ☐
More than 500 ☐

SECTION III: YOUR GENEBANK'S ORGANISATION

12. *How does your genebank fit into the structure of your parent institute?* **A** *N A L Y S I S*

Genebank is independent of other projects but sometimes works with other projects ☐
Genebank is independent of other projects but works closely with other projects ☐
Genebank is part of another project ☐
Other (please specify) ☐

_____

_____

13. *Does your genebank work on a collaborative basis with other research stations (e.g. for the purposes of regeneration or evaluation of accessions?)* **A** *N A L Y S I S*

Occasionally ☐
Frequently ☐
No, never ☐

If the answer to question 13 is "no, never", go straight to question 17, otherwise complete questions 14 & 15.

14. *How is the work organised at the different research stations?* **A** *N A L Y S I S*

By your genebank ☐
In consultation with your genebank ☐
Independently of your genebank ☐

15. *Do the different research stations operate, or plan to operate, documentation systems?* **A** *N A L Y S I S*

Yes ☐
No ☐

If the answer to question 15 is "no", go straight to question 17, otherwise complete question 16.

---

16. *Have these documentation systems been developed independently of your genebank?* **A** *N A L Y S I S*

Yes ☐
No ☐

---

17. *Which of the following statements most accurately describes the organisation of duties at your genebank?* **A** *N A L Y S I S*

People do not have clearly defined duties and share duties on a day to day basis ☐
People have clearly defined duties but there is a regular sharing of duties on a day to day basis ☐
People have clearly defined duties but perform other duties ***when necessary*** ☐
People have clearly defined duties with little or no sharing of duties ☐

---

18. *In which areas is there usually a sharing of duties?* **A** *N A L Y S I S*

Indicate the approximate number of people involved

		Number:
Germplasm collecting	☐	_____
Acquisition of new germplasm samples	☐	_____
Germplasm health/quarantine	☐	_____
Sample registration	☐	_____
Seed cleaning	☐	_____
Seed drying	☐	_____
Seed moisture content testing	☐	_____
Seed viability testing	☐	_____
Seed packing and storage	☐	_____
Distribution of germplasm	☐	_____
Multiplication/regeneration of germplasm	☐	_____
Dissemination of information on germplasm	☐	_____
Selection of germplasm for distribution	☐	_____
Characterisation and evaluation of germplasm	☐	_____
Research (e.g. germplasm enhancement, seed physiology)	☐	_____
Organisation of technical meetings/training workshops	☐	_____

19. *In which areas is there usually backlog of documentation work?* **ANALYSIS**

Indicate the extent (e.g. small, medium, large)

	Extent:
Germplasm collecting	☐ _____
Acquisition of new germplasm samples	☐ _____
Germplasm health/quarantine	☐ _____
Sample registration	☐ _____
Seed cleaning	☐ _____
Seed drying	☐ _____
Seed moisture content testing	☐ _____
Seed viability testing	☐ _____
Seed packing and storage	☐ _____
Distribution of germplasm	☐ _____
Multiplication/regeneration of germplasm	☐ _____
Dissemination of information on germplasm	☐ _____
Selection of germplasm for distribution	☐ _____
Characterisation and evaluation of germplasm	☐ _____
Research (e.g. germplasm enhancement, seed physiology)	☐ _____

20. *What are the genebank's documentation priorities?* **ANALYSIS**

Indicate the degree (e.g. low, medium, high)

	Degree:
Germplasm collecting	☐ _____
Acquisition of new germplasm samples	☐ _____
Germplasm health/quarantine	☐ _____
Sample registration	☐ _____
Seed cleaning	☐ _____
Seed drying	☐ _____
Seed moisture content testing	☐ _____
Seed viability testing	☐ _____
Seed packing and storage	☐ _____
Distribution of germplasm	☐ _____
Multiplication/regeneration of germplasm	☐ _____
Dissemination of information on germplasm	☐ _____
Selection of germplasm for distribution	☐ _____
Characterisation and evaluation of germplasm	☐ _____
Research (e.g. germplasm enhancement, seed physiology)	☐ _____
Organisation of technical meetings/training workshops	☐ _____

5 Analysis of the questionnaire

Now that you have completed the questionnaire, you are in a position to analyse your responses to the different questions. Work through the following analysis section, referring to the responses you have given to each question. Any conclusions you make or ideas you have, should be noted in the right hand area of your form, headed "Analysis". This will provide you with a concise record of your responses to these important questions about your genebank, an interpretation of your responses and suggestions for how they can be used in the design of your documentation system.

You will also find it helpful, when you reach Chapter 4, to transfer any comments that refer to specific genebank procedures to the forms that are supplied there for the analysis of procedures.

ANALYSIS - SECTION I: YOUR GENEBANK'S RELATION WITH OTHER GENETIC RESOURCES PROGRAMMES

1. *Which category does your genebank fit into?*

The answer to this question will help determine the size and sophistication of the documentation system needed and identify whether you need to consider data exchange facilities. National, regional and international genebanks commonly operate on a larger scale than institutional genebanks. They therefore have a greater need for computerized documentation systems capable of data exchange with other plant genetic resources programmes.

2. *When was your genebank established?*

Generally speaking, the older the genebank, the more established the activities and procedures. Procedures which are firmly established are generally easier to document than those which are under development. Therefore, if the genebank is newly established or there are areas under development, greater care needs to be taken when analysing procedures and developing the documentation system.

3. *Is there a national programme* of *plant genetic resources activities in your country?*

4. *Does your genebank collaborate with other genetic resource programmes in other institutes* or *genebanks?*

If the answer to either of these questions is "yes", you should investigate the need for data exchange and the adoption of common data formats by the different collaborators. Think about standardising descriptor lists, coding systems, software and system structure. Find out how other genebanks manage their documentation. Do they currently run manual or computerized systems?

ANALYSIS - SECTION II: YOUR GENEBANK'S AIM AND AREAS OF ACTIVITIES

5. *Summarise in a short phrase your genebank's aim*

6. *Which* of *the following activities does your genebank undertake?*

The answers to these questions will tell you about the main focus and direction of the genebank's work and the activities which have to be addressed when designing the documentation system.

7. *In which form(s) is/are the germplasm maintained and in which type* of *collection?*

Each collection of germplasm will be managed quite separately and have different documentation requirements to the other collections. Each collection of germplasm should therefore be analysed separately and treated as a separate section of the documentation system (e.g. by the use of separate files, procedures etc.).

8. *How many accessions are maintained by your genebank?*

9. *How many different species are held?*

10. *Is your genebank in an expanding* phase?

The answers to these questions will help you determine the need for a computerized documentation system. In general terms, when a genebank has between 100-500 accessions, a computerized system is

desirable; between 500-1,000, it is highly desirable; more than 1,000, it is essential.

The number of species also needs to be taken into account when making this assessment. Collections maintaining single species or crops are the most straightforward to manage and document. Where there are several species, these should be analysed separately and be treated differently in the documentation system as appropriate. Therefore, the greater the number of species, the greater the potential complexity of the system, so the greater the need for a computerized system.

If the genebank is in an expanding phase, the documentation system should be flexible enough to cope with the future changes. An increasing workload suggests a greater need for a computerized system.

11. *Approximately how many samples do you distribute in a typical year?*

Genebanks need to keep, track of where germplasm samples are being sent. When this number is fewer than 100 per year, a manual system or a simple inventory manager is usually quite adequate. However, when this number is larger, a more sophisticated inventory manager is needed (usually computerized) to keep details of recipients, distribution dates etc., for future analysis.

ANALYSIS - SECTION III: YOUR GENEBANK'S ORGANISAIION

12. *How does your genebank fit into the structure of your parent institute?*

The purpose of questions 12-16 is to determine the possible operation and organisation of the genebank documentation system.

Your genebank might be part of a larger organisation where there is a documentation effort-in existence. If this is true, your documentation system may have to comply with existing documentation standards. Alternatively, your genebank may have close links with other projects or departments whose activities require documentation in your documentation system.

If your genebank is completely independent, then a documentation system totally specific your needs can be developed.

13. *Does your genebank work on a collaborative basis with other research stations? (e.g. for the purposes of regeneration or evaluation of accessions).*

If the answer to this question is "yes", you should investigate more thoroughly the mechanisms of data exchange and the possibility of operating a similar documentation system at each research station.

14. *How is the work organised on the different research stations?*

If work at other sites is organised by your genebank, you should look closely at the different activities taking place. Assess the documentation and information needs, of the other research stations, paying particular attention to information that is needed to plan activities and manage resources.

15. *Do the different research stations operate, or plan to operate, a documentation system?*

If the answer to this question is "no", then you will probably operate a single documentation system based at your genebank.

16. *Is the documentation system being developed independently of your genebank?*

Where possible, common standards should be used in the design of the different systems. This includes the same descriptor lists, coding systems, software and basic system structure. Common standards will help in the future exchange of data between the research stations.

If answer to this question is "no", the Dossibility of operating a similar documentation system at each research station should be investigated. If the answer to this question is "yes", do not feel that the design of your documentation system needs to be identical to that of other existing or planned systems. However, the ability to exchange data should be made a priority and allowed for in any system design: First of all, determine your own documentation and information needs and then (and only then), assess the other systems to determine whether they are suitable for your requirements.

17. *Which of the following statements most accurately describes the organisation of duties at your genebank?*

In general, if people have clearly defined duties which are not shared, it will be easier for you to design, implement and operate documentation procedures. Responsibilities for performing specific procedures can be allocated more easily. It will also make the process of assessing your documentation and information needs much more straightforward.

18. *In which areas is there usually a sharing of duties?*

19. *In which areas is there usually a backlog of documentation work?*

20. *What are the genebank's documentation priorities?*

Where there is a sharing of duties, the design, implementation and operation of documentation-procedures must be studied in great detail. For instance, if you designed a manual form for recording data, it would cause problems if several people needed to use it at the same time. It is also essential to develop a standard method for performing each documentation procedure and to ensure that anyone who will be involved in the procedure is aware of this. We'll return to this subject in Chapters 6 and 10.

Indicating a relative priority for the documentation of each of the different procedures will help you to identify the actual need for documentation. High priority procedures will also have greater influence on how the system is structured. Prioritising will help you plan the implementation of the new system once it has been designed. Don't worry at this stage if you don't know the relative priorities of different procedures, as this will become clearer when we start to look at the genebank procedures themselves in Chapter 4.

Identifying areas where there is a backlog in documentation will alert you to procedures where large amounts of data are generated, existing documentation procedures are time-consuming and/or responsibility for documentation of a particular procedure has not been allocated wisely. Establishing the cause of any backlog will indicate areas you need to give careful thought to when planning the implementation of your documentation system. Also, you will save yourself a lot of time if you implement documentation procedures *before* you clear any backlog of work. So, when we start to analyse the genebank procedures in Chapter 4, look carefully at the areas you have indicated to discover the scale of the backlog and when it is likely to be cleared.

Chapter **3**

Information processing in genebanks

Chapter 3 highlights the use of accession numbers, batch references and scientific names for relating data together to build the genebank documentation system and describes the central role played by genebank procedures in the generation and processing of information. When you have finished this chapter you will be able to:

- Define accession-specific data and group data
- Discuss the use of accession numbers, batch references and scientific names in relating data together
- Explain why procedures are operated
- Classify genebank procedures as operational or scientific
- Discuss the handling of comments in documentation systems
- Discuss the design considerations when constructing a flow chart

1 Relating data together from different activities

The data that are generated and used in a wide variety of genebank activities will fall into two main categories: accession-specific data and group data.

Accession-specific data concern individual genebank accessions. Most of the descriptors you use will be for accession-specific data such as seed moisture content, % viability, weight of seed, 1000 seed weight, plant height, seed colour and so on.

Group data concern groups of accessions. For instance, you'll find data referring to particular species (e.g. equilibrium moisture contents, viability test methods, regeneration procedures, methods for determination of moisture contents, etc.). You'll also find data relevant to many genera and species such as parameters for modified sequential germination testing.

These categories of data will be handled quite differently in your documentation system. In this guidebook, we'll be principally concerned with building a documentation system for accession-specific data since these data are important for genebank management.

Accession numbers, batch references and scientific names are used as the basis for building up the documentation system

How can you relate together and co-ordinate all the different accession-specific data in your documentation system? You need to use the accession numbers, their batch references (where appropriate) and their scientific names as the basis for building up the documentation system. As these are fundamental to the management of the genebank and the documentation system, it's worth examining them more closely.

1.1 Importance of the accession numbering system

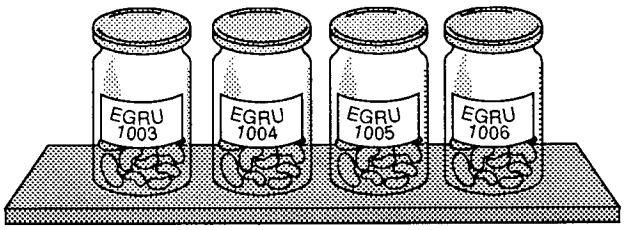
Each genebank accession will have its own, unique accession number which distinguishes it from all other accessions in the genebank. If an inappropriate numbering system is used or if none is used at all, the consequences are serious: it will cause confusion, create much unnecessary work through mistakes being made and ultimately lead to loss of valuable germplasm.

If you don't already have an accession numbering system in operation at your genebank, what sort of system should you use to facilitate management of the genebank and the operation of the documentation system? You need to use a system which is *simple* and *practical* to use. Remember that you'll be writing accession numbers in several places, whether on forms or in reports; or entering the numbers in the computer. If the accession numbers are long or complicated, it makes their recording much more prone to error. If errors are made recording the accession numbers, the consequences can be disastrous. Therefore, *keep the numbering system simple!*

The simplest and most practical approach is discussed in the following sections.

1.1.1 Use a strictly numeric system which is sequential in operation

It should start at "1" and then increase as each new accession is received i.e. "2", "3", "4" and so on. So that you don't confuse the accession numbers with any other numbers, you'll add your genebank acronym to the number. For example, if the acronym of your genebank is "EGRU", the accessions will be labelled EGRU 1, EGRU 2, EGRU 3 and so on.



1.1.2 Do not incorporate additional information in the accession number

The accession number should be used for uniquely identifying the accession and nothing else. Don't try to incorporate additional information such as the year of deposit or a crop code into the number – it might cause confusion later on.

1.1.3 Use only one numbering system

Operate a single numbering system which is used for ALL accessions.

Since the accession numbering system is such an important concept for managing the genebank and operating the documentation system, it's worth looking at the consequences of using inappropriate numbering systems.

1.1.4 Do not use the depositor's designation

If you only have a small genebank, you have might considered using the depositor's designation as the basis for your accession numbering system rather than devising your own system. This approach can cause serious operational difficulties in your genebank and *should not be attempted.*

Example

Suppose you receive an accession with the number 240ß-1D. In the next month you might receive a number of accessions labelled: ZP414, 240-2, Três Marias, C3P-0, Σ 240, Parson Thirdly, 240α , RB12. However, during the genebank accession procedures the technician can't quite read the writing on the seed packets so Σ240 is written as E240, 240ß-1D as 240B-lD, Três Marias as Tras Manas, 240α as 240a and so on. The mistakes aren't noticed until much later when you're looking in the cold room for accession "Três Marias". You've been looking round for sometime but all you can find is something called "Tras Manas". Is it the same accession? You spend another hour going through the records and asking the technician. You come to the conclusion that it is the same accession. Well it *probably* is...

As illustrated in this example, mistakes can easily be made in writing the depositor designations and the mistakes are not easily noticed. This inevitably leads to confusion and the creation of additional and unnecessary work.

1.1.5 Do not use the germplasm collector's number

If your genebank is quite small but all your accessions are received from collecting missions you might consider using the collector's numbers as accession numbers. As in the previous section this approach *should not be attempted.*

Example

You might receive a series of accessions from a collector which are labelled LG1, LG2, LG3 and so on. The next batch of accessions are from a different collector and they're labelled AB22, AB23, AB24 and so on. So far, so good. Six months later, the first collector deposits some more germplasm samples labelled LG1, LG2, LG3 and so on. They're not the same germplasm samples as the first batch as they were collected from a different region. What do you do now?

1.1.6 Do not use a different numbering system for each crop

Genebanks which maintain very large collections of germplasm sometimes operate separate but sequential accession numbering systems for each crop. If you maintain many different crops or only work in a small to medium sized genebank, this approach is *not recommended.*

Example

Suppose your genebank maintains about a thousand accessions of a few crops, say *Capsicum, Lycopersicon, Hordeum, Arachis, Vicia* and *Pinus* and that each crop has its own sequential and numeric accession numbering system. So for *Capsicum* the numbering system is CA1, CA2, CA3 . . . for *Arachis* it's AR1, AR2, AR3 and so on. This system seems to work for several months. Then the curator decides to accept a collection of pigeon pea, *Cajanus.* What numbering system do you use for this? You can't use "CA" because that's already being used for *Capsicum.* So you might decide to use "CAJ" instead. A few months later, the genebank manager decides to accept a modest sized collection of forage legumes comprising many different genera and species, perhaps only 10 of each species. You suddenly realise that you-have 20 or more separate numbering systems in operation.

This situation is clearly undesirable. In most cases it's far simpler to operate a single, sequential and numeric accession numbering system.

1.1.7 Do not use a "reserved" numbering system

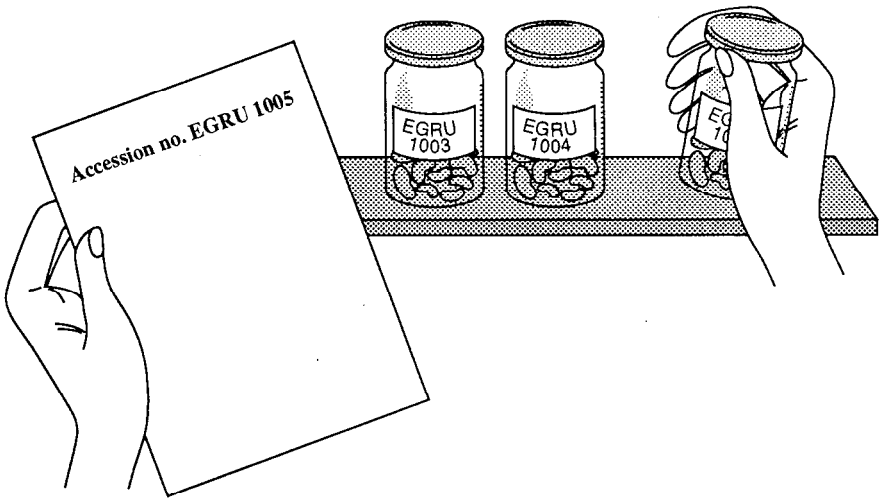
You might consider using a single accession numbering system but "reserve" numbers for particular crops, for instance 1 to 100 for barley, 101 to 200 for wheat, 201-300 for Triticale and so on. This system can cause a number of operational difficulties in the long run and *should not be attempted.*

Example

Using the reserved numbers (1 to 100 for barley, 101 to 200 for wheat, 201-300 for Triticale and so on), consider what would happen if you reached 100 for the barley accessions. You might reserve numbers 501 to 600 for further barley accessions. The same situation might arise for other crops. After a few years you may have some 40 "reserved" numbers in the genebank and nobody (not even yourself) can remember what the different reserved numbers actually refer to. Also you can't help wondering whether the gaps in the seed store are because there are some numbers yet to be allocated or whether some seed packets are missing...

Clearly, if you are going to use a single, numeric accession numbering system, it's so much easier if it's *strictly sequential:* it starts at "1" and continues.

From these examples you can see that the accession number is a unique identifier which distinguishes one accession from another and you can see the advantages of having a single accession numbering system which is numerical and sequential. The only occasion when more than one numbering system is justified is when the purpose of the collection is different from the main collection, such as a temporary in-house collection or a reserved and "restricted distribution" collection. If most of the in-house collection end up in the main genebank collection then even this in-house numbering system is unnecessary. Circumstances will tell you if such separate numbering systems are necessary.



NEVER re-use an accession number

Remember that any accession numbering system should be operated carefully to prevent any errors from occurring. An important rule which should not be broken is that an accession number should *never* be reused, even if an accession is lost for any reason or dies on storage. Even though the accession is lost, there will still be information about it in the documentation system and in printed reports, catalogues, scientific publications. Also, you might have distributed the accession to outside users

– just think of the confusion that would be caused if you subsequently used the accession number for a different accession. So, keep it simple. Give each new accession a new accession number.

1.2 Importance of the batch reference

Definition BATCH REFERENCE

Any date, code or number which uniquely identifies the regeneration cycle of the accession.

Examples of batch references are given in table 1.

Table 1. Different ways of expressing the batch reference

BATCH REFERENCE	EXAMPLES	COMMENTS
Date	16-Feb-1990	The precise date on which regeneration was started (sowing date, germination date, etc.)
Month and year	Feb-1990	The month and year in which regeneration was started. Not precise enough for *in vitro* subculturing which can be in frequent intervals
Season and year	Spring 1990 Winter 1990	The season and year in which regeneration was performed. Not precise enough for *in vitro* subculturing
Field number and year	1028-l990, 1132-90	The reference to where the accession was regenerated in the field together with the year of regeneration. Inappropriate for *in vitro* subculturing
Number of the regeneration cycle	6, 12	The 6th or 12th time the accession has been regenerated. A drawback is that there is no indication when the regeneration took place
Pedigree	1.3.4.1 1.1.2.1	Used in *in vitro* collections to show the pedigree of the batch. This is discussed in chapter 4, section 4

When handling and documenting genebank accessions, it is common practice to record the batch reference of each accession along with any other data. Why is it so important to record it so routinely?

When a genebank receives a sample of germplasm, say from a collecting mission, the accession is not genetically uniform – it contains a certain amount of variability which might be detected in a characterisation trial. Each time the accession is regenerated there is a chance that the

natural diversity of the accession will be affected (see fig. 1). It's important therefore that whenever an accession is used for *any* purpose, a record is kept of which batch reference is used so that any data generated can be related to specific batch references.

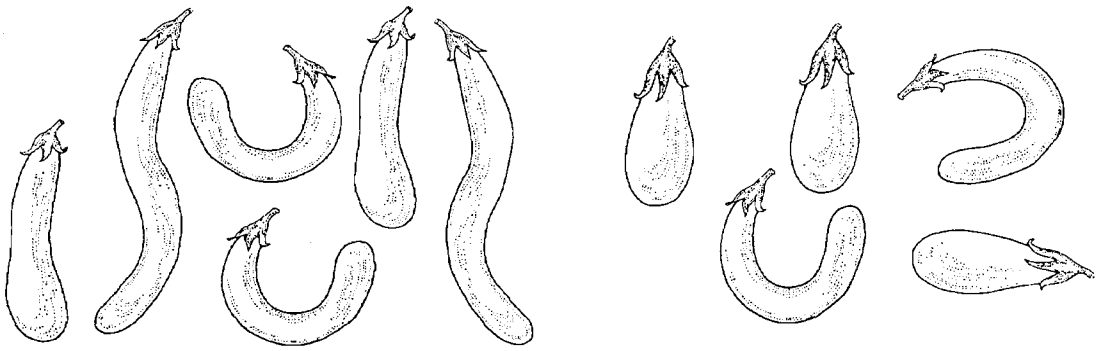


Fig. 1. Two different batches of the same accession of Solanum melongena (aubergine) that show variation in fruit shape and size

For instance, if there were any differences between characterisation trials using the same accessions, these differences might be because:

a. Different characterisation conditions were used
b. The trials used different batch references of the same accessions
C. A combination of these effects

When you documented the different trials, you wouldn't mix data from the different trials; you'd document them separately for the sake of accuracy, recording the batch references for all the accessions used. Similarly, if you were distributing accessions, you'd want to know precisely which batch references of the accessions were being distributed in case there was a query some time in the future.

How do you choose which sort of batch reference to use – season and year, sowing date, field number and year? Remember that it should *uniquely* identify the regeneration cycle – it should not be possible to confuse it with any other regenerations of the same accession. So, for *in vitro* work, the month and year as a batch reference would be inappropriate if you were subculturing, say, every 3 weeks. You will also want to choose one which is straightforward and easy to record. For many people, a batch reference based on a date fulfills these requirements. It has the added advantage of indicating the approximate age of a sample which can be useful in many management activities.

1.3 Importance of the scientific name

When you publish any of your genebank data in papers, reports or newsletters or exchange germplasm and information on germplasm, you will have to use the scientific name instead of (or in addition to) the crop name. The scientific name is an internationally recognised Latin name for the crop and facilitates the collation of information on a species from many disciplines. The crop name by itself is not usually sufficient as it varies so much from country to country. For instance, the following crop names all refer to the same crop, *Secale cereale:* rye, centeio, Roggen, centeno, seigle, barakka, çavdar (see fig. 2). It's usually more convenient to use the crop name for every day use but you need the up-to-date scientific name for all publications and communication with other genebanks and genebank users. Often you will have to quote the authority for the species (e.g. *Secale cereale* L.). This is particularly important for those taxonomic groups where there are re-classifications and changes in nomenclature. For this reason, many genebanks routinely record this information in their documentation system.

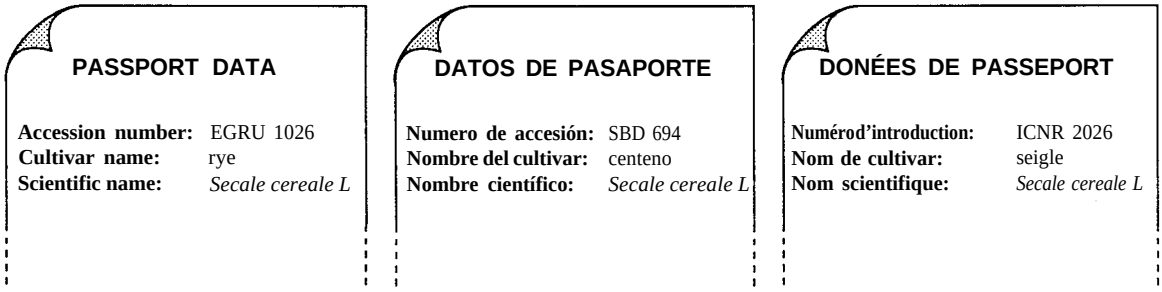


Fig. 2. The scientific name can identify a crop when working in different languages (as illustrated with English, Spanish and French in this example)

2 Introduction to procedures

Genebank procedures are at the heart of data generation and processing. It's for this reason that all genebank procedures should be studied carefully for:

- Data produced
- Data used
- The management value of data produced
- Information requirements from the documented data
- The need and priority for documenting data

Before we look at data generation and processing, it's worth looking in general terms at procedures so that you can understand why they are performed and, at a later stage, develop procedures yourself that will be used in the documentation system.

2.1 What is a procedure?

A procedure is a sequence of actions which collectively accomplish a desired task. You probably have several procedures in your genebank which are routinely followed. These might include seed-handling procedures (e.g. sample registration, seed drying, viability testing, seed packing and seed distribution) and documentation procedures (documentation of characterisation data, production of reports). All of these procedures consist of a series of step-by-step instructions which, when followed correctly, perform a desired task.

It should be emphasised that if it's not clear what is being achieved in the procedure (and why the procedure is being performed in the first place), the chances are that mistakes will be made performing the procedure and that documentation of data generated and used will be difficult. It's as simple as that!

2.2 Procedures simplify working practices

Procedures are set up:

- To perform a specific task
- To simplify the task for the person performing it and the person managing it.

It's much easier if people follow the same procedures in the genebank as it reduces any confusion and is an effective use of available resources. Just imagine the confusion that might arise in a large seed handling unit if everyone had their own special way of carrying out seed viability tests: would the results be reliable? Would everyone make optimum use of their time and the equipment available? Probably not!

2.3 Availability of human and physical resources

When designing a procedure, careful thought is given to what is being achieved and how best to perform the task with the human and physical resources available. There's more than one way that the task can be performed and therefore more than one procedure that can be designed to do it – not all of these will be practicable however – a lot depends on the resources in the genebank. It follows also that procedures may differ between genebanks, each procedure being designed according to the genebank's needs.

2.4 Badly designed procedures create more work

If a procedure has been badly designed or is missing altogether, there can be a lot of confusion and mistakes can easily be made. This will inevitably create more work in the genebank – correcting mistakes which should have been avoided in the first place. You can imagine the mistakes that might be made if there were no clear procedures for seed drying and seed moisture content determination – the tasks would be difficult to perform and valuable seed would probably be lost.

3 Data generation and use in procedures

3.1 Classification of procedures

It's useful to classify genebank procedures into two major classes: operational and scientific.

Operational procedures are concerned with the day to day operation of the genebank. These include: sample registration, seed cleaning, seed moisture content determination, seed drying, seed viability testing, seed packaging, seed storage, seed monitoring and seed distribution.

The data produced in these procedures are of vital importance for the management of genebank's collections and therefore the continued viability of the germplasm. As these data are so important, a *high priority* is usually given for their documentation.

Data from operational procedures tend to be dynamic, that is, they need to be updated at a later date. For instance, seed viability will decline on prolonged storage; the amount of seed in store will decrease over time as seed is used in the genebank or sent to other scientists. If the information retrieved from the documentation system is to be of any use, it's important that these data are kept up to date.

Scientific procedures generate data which are of potential interest to those working outside the genebank. Examples of scientific procedures include anything to do with germplasm characterisation and preliminary evaluation. Unlike operational procedures, information from scientific procedures tend to be stable. For instance, characterisation data once recorded are unlikely to change, say, after a year or so. Since the information is of interest to others, you'll have to think of ways of communicating this information to other scientists using your documentation system.

Examples of **scientific procedures** include: germplasm characterisation and preliminary evaluation

In most genebanks, a higher priority is given for the documentation of operational procedures than for scientific procedures, as data from scientific procedures are not used for the day-to-day management of the

genebank. This does not mean that *all data* produced in operational procedures need to be documented. Nor does it mean that scientific procedures do not need to be documented at all.

We can summarise some of the features of operational and scientific procedures in table 2. Some of the features are discussed in greater detail in later sections.

3.2 Relationship between procedures

Table 2. Features of operational and scientific procedures

	RELATIVE MANAGEMENT VALUE OF DATA PRODUCED	RELATIVE PRIORITY FOR DOCUMENTATION	RELATION TO OTHER PROCEDURES	UPDATING OF DATA REQUIRED?
Operational procedures	High	High	Can form part of a chain	Yes
Scientific procedures	Low	Low	Usually stand-alone	Not usually

Scientific procedures are usually "stand-alone". **Operational procedures** are usually dependent on, (or related to), other genebank procedures

Scientific procedures are usually "stand-alone", that is, they can be performed independently of other procedures. A good example is a characterisation trial.

By contrast, operational procedures are usually dependent on, (or related to), other genebank procedures. Such procedures form part of a chain and are always performed in sequence. An example of a chain of related operational procedures is illustrated in fig. 3.

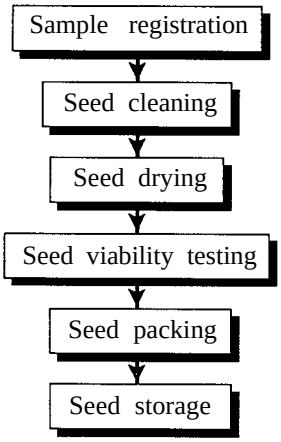


Fig. 3. Operational procedures performed in a specified sequence

In this example, seed packing would not be carried out unless the seeds had been dried and found to be of a sufficient viability to store.

Operational procedures often form part of a chain of procedures but can operate independently, e.g. viability testing, regeneration, seed distribution. For example, seed viability testing is a routine part of the overall accessioning of a sample as well as a procedure used for monitoring batches in the seed store.

For the genebank procedures which are related you should construct a flow chart showing in which order they are performed. Flow charts are a useful way of getting an overall picture. Once the chart has been constructed, the information flow becomes clearer as well. We'll look at flow charts later in this chapter.

3.3 Time scale

You'll need to know how much data are being handled over a given period so that you can make decisions about the most appropriate way to document them. It may be that a procedure generates a lot of data but is only performed once a year (e.g. a characterisation trial). Alternatively a procedure may produce few data but is performed regularly (e.g. moisture content determinations).

3.4 The management value of data generated

When documenting operational procedures, the question that needs to be asked is "What is the management value of these data?" In order to answer this question you have to understand why the procedure is being performed. You also have to assess the need for the data to be documented. Remember that the documentation system will provide information at a later stage, usually in response to questions. If you have a clear understanding-of the purpose of the procedure and the sorts of questions that will be asked later on, it's much easier to identify those data which need documenting and the best way to document them.

For instance, a genebank procedure for seed viability testing is performed to obtain data on the seed viability of particular seed batches. Clearly, these data need to be documented so that future decisions can be made on regeneration priorities using information from the documentation system.

The difficulty arises where the procedure is not performed with the aim of generating data but data, nevertheless, need to be documented. A good example is seed distribution – the aim is to distribute seed but data need to be recorded on how much seed is sent, the date it is sent and to whom it is being sent. In the next chapter we'll be looking more closely

at the different genebank procedures and the information requirements of the data once documented.

3.5 The number of comments recorded

There will always be occasional observations that are made during genebank procedures which might be important (now or in the future) which you'll want to document but which don't fit into any of the groups that you define. These comments are particularly common in newly established or developing procedures. However, even the most established procedure should have room for these miscellaneous data.

Sometimes comments can be a source of useful information but they are a problem to handle effectively because they can be something of an unknown quantity – they need not refer to the same subject, they are not regularly made, they are not widely made across the same crop and they have no predictable structure. When comments are widely and regularly made on the same subject, there comes a point when they should be incorporated in the documentation structure as a separate item.

If many comments are recorded in the procedure you should ask yourself the following questions:

3.5.1 Is the objective of the procedure clearly defined?

Sometimes the reason why so many comments are made during the procedure is that the objective of the procedure is not focussed enough. In this case, the user will note down anything that comes to mind in an unstructured way. If the objective of the procedure is clearly defined then it's easier to identify the data that need documenting.

3.5.2 Is the procedure under development?

If the answer is "yes" it may be that the data generating areas are changing or that there are changes in the working practices.

3.5.3 Are the appropriate descriptors being used?

It may be that certain descriptors are missing or the descriptors used are inappropriate for the observations that are being made. For instance, the descriptor "Herbage yield at 3 weekly intervals" could not be used if measurements were made every 4 weeks.

3.6 Resource requirements

It's important to make the distinction between procedures which are labour intensive and those which are documentation intensive. Remember that some procedures might be lengthy or labour intensive but not actually produce much data which need documenting. Many operational procedures are like this (e.g. seed cleaning).

If data recording/analysis is a major use of time, the procedure is documentation intensive. You'll have to spend time making sure that the documentation procedure is an integral and efficient part of the overall procedure and that it assists, rather than hinders, the users. This is the case in scientific procedures such as characterisation trials where much field work is done; many observations are recorded in the field and then formally documented in the laboratory. Where more than one person is involved, the procedure will work most efficiently if there is a clear idea of each person's responsibilities and who is in charge. This too will help the documentation process if documentation is an integral part of the genebank procedure.

4 Construction of a flow chart

Flow charts are essentially a visual aid to help the user get an overall picture of the different steps of a procedure and the different relationships involved. During your genebank analysis, you may find it useful to construct a flow chart for each procedure showing the different steps of the procedure, the decisions made at each step and any data used/generated. You can also construct a flow chart or a table showing the relationship between different genebank procedures.

There are several ways to construct a flow chart. A flow chart for a procedure could be constructed as follows:

1. List each step of the procedure in sequence vertically down the page with the decisions (if any) made at each step
2. Next to this list the data/information used or generated

An example of this type of flow chart is given in fig. 4, section 4.2.

A flow chart showing the relationships between procedures can be constructed as follows:

1. Arrange related procedures in sequence vertically down the page and next to each procedure name, briefly list:

   (a) the goal of the procedure

   (b) the data/information used or generated

This type of chart can also take a tabular form, as illustrated in table 3, section 4.3.

4.1 Other considerations

Don't make the mistake of trying to incorporate all genebank procedures on a single flow chart – only include those procedures which are related. Make additional flow charts to show different relations.

Keep the flow chart simple – you should be able to make some sense of the chart when it is finished. Remember that you make flow charts to clarify the procedures and their relationships, not make them more complex or obscure.

Genebanks follow different procedures to accomplish essentially the same task. The following examples are not therefore supposed be directions on how to perform the procedures – make your own version to suit your genebank!

4.2 Example 1: Sample registration procedure for new accessions

This procedure can be described as follows:

"Seed samples for inclusion in the genebank are received from the depositor and placed in a temporary store. A check is made to ensure that the genebank does not already have these accessions from another source. If any of the samples already exist in the genebank, a decision is made on the fate of the sample (e.g. return the samples to the depositor). The samples for inclusion are arranged in order of depositor number and each given an accession number in a sequential fashion. These accession numbers are written on the seed packets as well as any documentation that was supplied with the seed packets.

Next, each packet of seed is opened to check that they are in good condition and that there are no damaged, broken, empty, shrivelled seeds or signs of infestation by fungi or insect pests. If the seeds are not in good condition, further tests are performed. If the seeds are in good condition, the registration procedure is continued and all passport data (including any characterisation data) documented using the genebank accession numbers."

The different steps can be summarised on a flow chart in fig. 4.

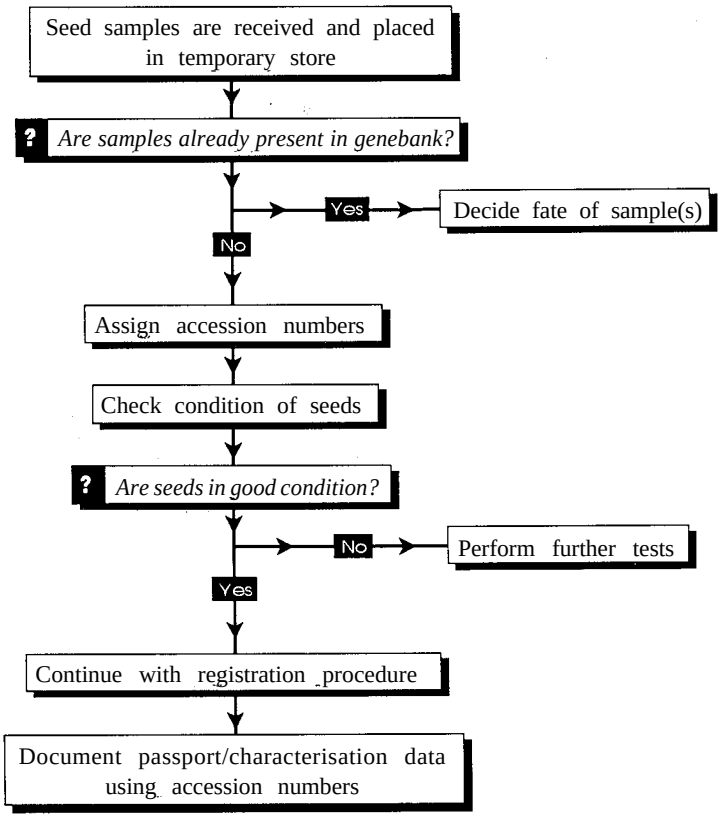


Fig. 4. Example of the steps in sample registration for new accessions

4.3 Example 2: Sequence of procedures performed on receipt of seed sample

Table 3 summarises a sequence of procedures in the order that they are performed. The decisions that are made at each step are not included on this summary chart.

Table 3. Sequence of procedures performed on receipt of seed sample

PROCEDURE	ACTIVITY	INFORMATION
Registration	Allocate accession number	Passport data Characterisation data (Other)
Seed cleaning	Clean seed, remove debris	Comments on sample health
Seed drying	Dry seed to acceptable moisture content	Moisture content (fresh, final) 1000 seed weight Drying method
Seed viability	Check that seed has high viability	Date of test % viability Method used Date of next test
Seed packing and storage	Securely pack dry seed and store to ensure long term viability	Weight of seed % viability Minimum weight of seed allowed Date of next viability test Moisture content Location in store

5 Where to go from here

Once you have satisfactorily completed the exercises given in the next section, you'll be in a position to start the detailed analysis of the procedures that are used in your genebank. This analysis is covered in the next chapter.

6 EXERCISES

Exercises

1. Indicate whether the following statements are true or false:
   a. Most accession-specific data are the same
   b. Group data are concerned with groups of accessions
   c. Group data are handled in the same way as accession-specific data
   d. The accession number uniquely identifies a crop
   e. The accession numbering system should be strictly numerical and sequential in operation
   f. A different accession numbering system should be used for each crop
   g. An accession number should never be re-used, even if an accession dies on storage
   h. The scientific name of a crop varies from country to country and should not be used
   i. The accession numbers, batch references and scientific names are used for relating data together and building the genebank documentation system

2. Distinguish between accession-specific data and group data. Give examples of each.

3. Why is a higher priority given for the documentation of accession-specific data over group data?

4. Review the different approaches to accession numbering systems. Which system is the simplest and most practical to operate?

5. Explain why it is necessary to record the batch reference routinely in the documentation system.

6. Indicate whether the following statements are true or false:
   a. A procedure is a series of actions which collectively accomplish a desired task
   b. All genebanks follow the same procedures for particular tasks
   c. Operational procedures usually generate data of a high management value
   d. Data from operational procedures tend to be stable and do not usually need to be updated
   e. Operational procedures can often form part of a chain of procedures
   f. Procedures which are labour intensive are also documentation intensive
   g. If a procedure generates many comments it is possible that the procedure is badly designed

EXERCISES

h. Flow charts are used for showing the different steps of a procedure and the relationship between different procedures
i. Flow charts should be constructed to contain as much information as possible

7. What are the benefits of operating a well-designed procedure?

8. Distinguish between operational procedures and scientific procedures. Give examples of each.

9. Why is a high priority given for the documentation of data produced from operational procedures?

10. Explain why flow charts are used and how you can design them.

Chapter 4

Analysis of data generation and use in genebanks

Chapter 4 will look in greater detail at many of the commonly performed genebank procedures with regard to data generation and use. When you have finished this chapter, you will be able to:

- Discuss the aims and activities of commonly performed genebank procedures
- Discuss the relation of each procedure to other procedures
- Identify the data produced and used in each procedure
- Describe ways of saving time in documentation for each procedure

1 Introduction

In this chapter we'll be looking more closely at commonly performed genebank procedures with regard to data generation and use. We'll be looking at the reasons for carrying out the procedures and the data that are produced or used in the procedures. We'll also look at how the different activities are related to one another.

It should be stressed at the outset that:

- *This is not a complete list of genebank procedures.* You may not do all the procedures listed here or you may do many more. It depends on what your genebank was set up to do.
- *This is not a practical guide to genebank procedures.* It is an outline of common genebank procedures concentrating on their documentation. Genebanks show variation in the way they perform procedures and the order in which they do them. This variation depends on a variety of factors including the genebank activities, the resources available, the priorities set and the decisions taken.

In the final section we'll look at how you can perform the analysis in your genebank.

2 Seed collections

In this section, we'll be looking at procedures which are commonly performed by genebanks handling seed collections. These procedures can be visualised in the form of a flow chart (see fig. 1). This is just one example of how the procedures can be organised.

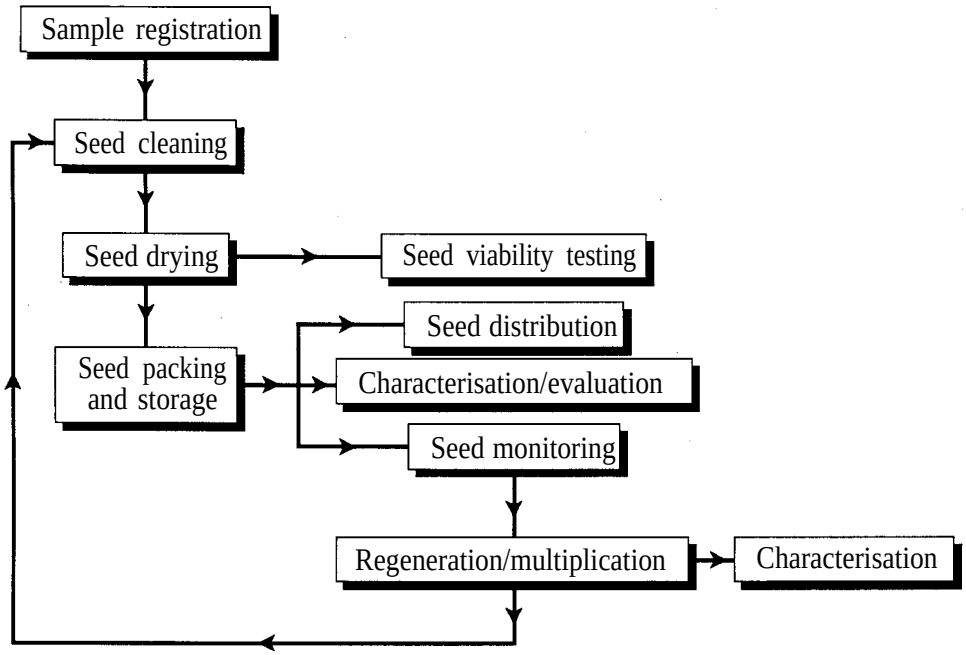
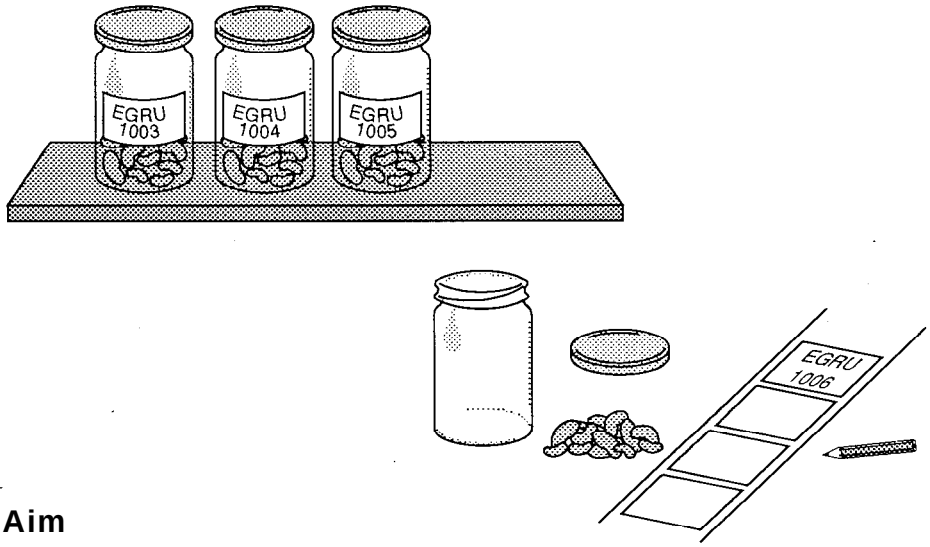


Fig. 1. Procedures commonly performed by genebanks handling seed collections

Using the flow chart in fig. 1, we'll be looking at each procedure with regard to:

- The aim
- The relation to other procedures
- The data produced and used
- The need for recording data in the documentation system: specifically the management value of any data and the information required from the documentation system

2.1 Sample registration



2.1.1 Aim

To register each new sample for inclusion into the genebank with a unique accession number and to record all relevant information received with the sample(s).

2.1.2 Outline

Seed samples are received for inclusion in the genebank. Unless the samples are already in the genebank, new accession numbers are assigned. Tests are performed on seed health before continuing with the accessioning process. If seeds are in good condition, data received with the samples. can be documented. All original documentation is filed for future reference.

2.1.3 General remarks

The primary aim of the registration procedure is to give each germplasm sample a unique accession number. Some genebanks prefer to do this after the seed cleaning procedure. In order to avoid the situation where the same accession number is accidentally assigned to more than one accession, it's a good idea to use a single file which is used to register the accessions (see table 1).

Sample registration is the first step in a chain of procedures which is followed for the majority of accessions (see fig. 1). A lot of work is involved in this chain of procedures and for this reason genebanks often adopt some sort of accession policy – a set of specific conditions which

have to be met if germplasm is to be accessioned. Rarely will genebanks accept just *any* germplasm!

2.1.4 Descriptor list for sample registration

An important part of the registration procedure is to document the data received with the sample. Much of this data will be *passport data*, that is, information recorded when the sample was originally collected, names or numbers allotted after collecting and any other relevant information. Passport descriptors can be conveniently divided into two categories, namely:

1. *Accession* descriptors – those concerning the registration of accessions at the genebank
2. *Collecting* descriptors – those concerning data recorded when the accession was originally collected

It's quite common to use a single file to document all the passport data particularly where only a few crops are maintained. Sometimes this is not always possible (or practical) particularly if a wide variety of collecting descriptors is being used for different crops. In such cases it is usually possible to use a single file to record accession data and separate files (one per crop) to record the collecting data (see fig. 2). Try to avoid the situation where more than one file can be used for accession data as mistakes in sample registration can cause serious problems later on.

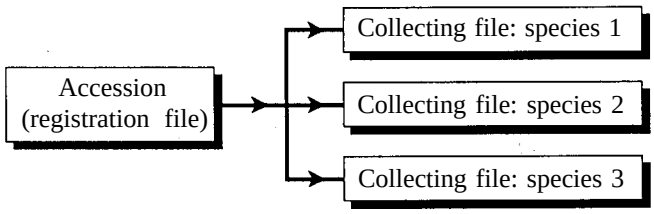


Fig. 2. A single registration file for all samples. Separate collecting files for each species

Some or all of the descriptors in tables 1 & 2 can be used in sample registration.

Table 1. Accession descriptors

DESCRIPTOR	COMMENTS
Accession number	
Scientific name	Genus and species is recorded. Sometimes necessary to record more detailed information such as subspecies and/or authority
Pedigree/cultivar name	Parentage, nomenclature and designations assigned to breeding material
Donor name	Name of institution or individual responsible for donating the germplasm
Donor identification number	Number assigned to the accession by the donor
Other numbers associated with the accession	Any other identification number known to exist in other genebanks for the accession
Acquisition date	Date on which the accession entered the collection
Date of last regeneration or multiplication	Either the date at the time of registration or the most recent regeneration date

Table 2. Collecting descriptors. Those marked with an asterisk* are descriptors which are considered essential

DESCRIPTOR	COMMENTS
Accession number	
Collecting number*	Original number assigned by collector of the sample
Collecting organisation*	Institute(s) collecting/sponsoring the sample collecting
Expedition identifier	
Collection data*	
Country of collecting*	
Province/state	Name of the administrative subdivision of the country in which the sample was collected
Location of collecting site*	
Latitude of collecting site*	
Altitude of collecting site*	

Table 2 (cont.). Collecting descriptors. Those marked with an asterisk* are descriptors which are considered essential

DESCRIPTOR	COMMENTS
Collecting source*	
Status of sample*	e.g. landrace, advanced cultivar, primitive variety
Type of sample*	Form in which the germplasm was collected (seed, vegetative, pollen etc.)
Scientific name*	
Local/vernacular name*	Name given by farmer to sample and the language or dialect used
Number of plants sampled*	
Land use	e.g. natural vegetation, grazing, urban area, etc.
Vegetation type	e.g. forest, grassland, bushland, etc.
Genetic erosion	Evidence for the threat of genetic erosion faced by the material at the collecting site
Plant use	
Cultural practices	
Photograph identification number(s)	Reference to photograph(s) taken at time of accessioning or environment at time of collection (if taken)
Herbarium identification number	Reference to herbarium specimen (if collected)
Topography	
Slope	
Aspect	
Annual rainfall	
Rainfall seasonality	
Soil type	
Soil pH	
Soil texture	

Passport data are not the only data that are documented at sample registration. There may also be data to document from the collector's observations of the accession's characteristics, or data from characterisation and evaluation trials performed after receipt of the samples (covered later in this chapter). A certain amount of miscellaneous data ("comments") might be supplied such as literature references, anecdotal data and ethnobotanical data (e.g. suitability for particular applications, unsubstantiated reports on performance in trials etc). We'll look at how you can handle these and any other feedback information in Chapter 6.

2.1.5 How to save yourself time in documentation

Sort samples into donor number order before allotting accession numbers

This might seem trivial but you'll be able to:

1. Spot any mistakes in your documentation later on
2. Produce easily readable reports or catalogues (which list both accession number and donor number)
3. Communicate more easily with the depositor later on

Mark all documents received with the new accession number(s)

Filing is so much easier if all documents received are marked with the accession numbers. Do this clearly, say, in red pen in the top right or left hand corner. e.g. "EGRU 196" for a single accession, or "EGRU 232-EGRU 244" for documents relating to a series of accessions.

Keep these documents in a file (or series of files) sorted in order of accession number. As your accession numbering system will be strictly a numerical and sequential one, the file will be a chronological record of the accessions received.

Insist that passport data are supplied with the samples

Part of the registration process is the documentation of information received with the samples. There's nothing so irritating as receiving the information 6 months or a year *after* you've accessioned the germplasm (or worse still, *never).* So, as part of the accession policy, why not insist that, at the very least, passport data are supplied with the germplasm?

Do not use abbreviations for scientific names

It is advisable to write the scientific name out in full and to avoid abbreviating it in the documentation system. This will avoid any pos-

sible confusion. For instance, it's not clear what *"P."* stands for – is it *Phaseolus, Physalis, Pouteria, Prunus* ...? Therefore, write the scientific name out in full.

2.2 Seed cleaning

2.2.1 Aim

To remove any broken or foreign seeds, debris, infested or infected seed from the sample.

2.2.2 Outline

There are a number of different procedures which are used for cleaning seeds depending on the species and the condition of the sample. The procedures vary in operation and in how labour intensive they are but have the following four basic elements:

- Initial examination
- Removal of debris
- Examination for fungal/insect infestation
- Removal of damaged/empty seeds

Seed cleaning will include a drying step for seeds with a high moisture content (e.g. fruits or recently harvested seeds). These would otherwise be damaged by the cleaning procedure. In this case, the sequence of events could follow the stages illustrated in fig. 3.

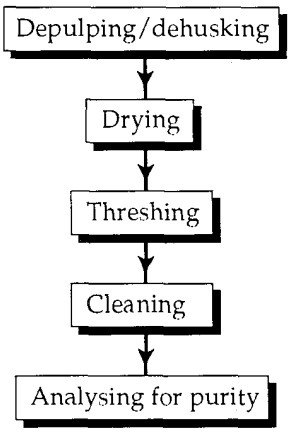


Fig. 3. Seed cleaning procedure for seeds with high moisture content

2.2.3 General remarks

Seed cleaning is a routine procedure and some genebanks choose not to document it except perhaps for the date of seed cleaning. However, this isn't always the case. Genebanks often set limits to the minimum number of seed they accept for accession; it can vary from 1,000 for a homogeneous population sample to 12,000 for a heterogenous population sample. Any fewer seed and the accession might be refused or multiplication be might necessary. In such genebanks, important data are recorded in the seed cleaning process, namely the estimate of the total seed number and the proportion of empty seeds. Also, if the seeds are treated in any way (e.g. with fungicides or by fumigation) this is also recorded.

2.2.4 Descriptor list for seed cleaning

Some or all of the descriptors in table 3 can be used in seed cleaning.

Table 3. List of possible descriptors for seed cleaning

DESCRIPTOR	COMMENTS
Accession number	
Batch reference	
Date seed cleaned	
Reference to method	Reference to standard method used in the genebank
Estimate of total seed	
Proportion of empty seed	
Treatment of seed	Name of chemical used
Operator	Name of person cleaning seed

2.3 Seed drying

2.3.1 Aim

To reduce the seed moisture content to accepted levels (without adversely affecting seed viability) and thereby increase seed life in storage.

2.3.2 Outline

The drying period of the seeds is predicted from estimates of their initial moisture content. The seeds are prepared for drying and subsequently dried. The moisture content is determined again; if it is not low enough, the seeds are dried for a longer period.

Subsequently the total weight of dried seed is measured along with the weight of 1,000 seeds (or 100 seeds for large seeds). Some genebanks choose to record the volume of 1000 seeds if they handle seeds on a daily basis by volume rather than by weight. Where large seeds are being handled (such as many beans and some tree species), genebanks commonly choose not to weigh the seeds but instead record the total number of seeds.

2.3.3 Descriptor list for seed drying

There are a number of different methods for determining or estimating the moisture content of seeds. Similarly there are a number of different methods for drying seeds. The methods you use will depend on the seeds you are drying and the equipment you have. A descriptor list for seed drying is given in table 4.

Table 4. Descriptor list for seed drying

DESCRIPTOR	COMMENTS
Accession number	
Batch reference	
Reference to method	Reference to method for seed drying/seed moisture content determination
Final moisture content determination (%)	
Date of moisture content determination	
Total dry weight of seeds	
1000 seed weight	
100 seed weight (for large seeds)	

2.4 Seed viability testing (germination test)

2.4.1 Aim

To determine, using a suitable germination test, what proportion of the seeds of an accession are alive and capable of producing plants under appropriate conditions.

2.4.2 Outline

This procedure is performed routinely after seed drying and periodically afterwards to monitor the viability of the seeds.

There is no general method for testing seed viability which can be used for all species; different species have different requirements for light, media, and temperature. However, there are three methods which-are most commonly used to determine seed viability: on absorbent paper, between absorbent paper and in sand. These methods have three steps in common:

- Preparation of seeds – commonly, two or more replicates for each batch are used
- Performance of test
- Interpretation of results – if the result is low, the test is repeated. A decision is subsequently taken on the fate of the batch

2.4.3 Descriptor list for seed viability testing

The aim of this procedure is very specific – it's to get data on the viability of the seed. These data are important and have a high management value. You'll have to think about which other data need to be recorded. Do you need detailed information about how the test was performed? Do you need a reference to the method used? You won't need details of the method if you always use the same one. A descriptor list for seed viability testing is given in table 5.

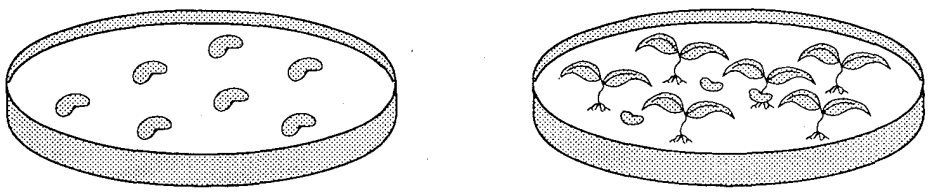


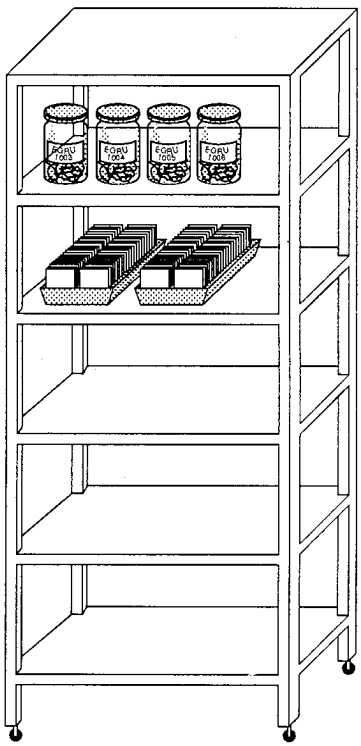
Table 5. Descriptor list for seed viability testing

DESCRIPTOR	COMMENTS
Accession number	
Batch reference	
Collection type	The collection from which the sample was derived, e.g. base, active
Reference to method	Reference to method used for seed viability testing
Date of viability test	
Viability (%)	
Operator	Name of person performing test

2.5 Seed packing and storage

2.5.1 Aim

To package seeds in such a way to prevent absorption of water from the atmosphere and contamination by pests; to store the seeds under conditions which maintain the viability and integrity of the accession.



2.5.2 Outline

- Prepare packaging. Seed packets or containers are prepared and labelled appropriately
- Package seed. A known amount of seed is placed in the packet/container and sealed. The seals on the packet/container are checked to make sure that they are intact and there are no signs of any other damage
- Determine location in seed store for seed packet/container
- Place packet/container in seed store
- Record location details in documentation system

2.5.3 Descriptor list for seed storage

Seed packing and storage form part of a chain of procedures and usually follow satisfactory seed drying and viability testing. Generally, seeds are only stored if they have an acceptably low moisture content and high viability level. As with registration, documentation is an integral part of these procedures and several important management data have to be recorded. A descriptor list for a seed store inventory file is given in table 6.

Table 6. Descriptor list: seed store inventory file

DESCRIPTOR	COMMENTS
Accessilon number	
Batch reference	
Collection type	The collection from which the sample was derived, e.g. base, active
Location in seed store	
Total amount of seed	
1000 seed weight	
100 seed weight (for large seeds)	
Minimum amount of seed allowed	Minimum amount of seed allowed before regeneration is necessary
Number of packets/containers	

2.5.4 How to save time in documentation

There are considerations which although not strictly the concern of the documentation specialist can help with the management of the procedure and the operation of documentation system.

Label seed packets/containers carefully with useful information

You should at the very least label the packets with the accession number and batch reference. However, you dont want to keep running from the cold room to the documentation system to find out what a seed packet is so you could consider labelling the packets with the following additional information.

- *Crop name.* It's useful if you put the crop name (or scientific name) on the packet, particularly if the seed store is not organised in accession number order but, say, by crop. You (or other genebank staff) will easily be able to spot seed packets which are in the wrong place.
- *Seed store location code.* If the seed store is not organised in accession number order, the location code will tell you what the "address" of a seed packet is. You'll be able to notice quickly whether a seed packet is in the wrong place.
- *Date of storage.* You'd use this information in making a decision on when next to check the seed viability. This is useful if you have seed packets from the same batch which were stored on different dates (rare in practice).
- *The packet number*. If you are storing the batch in more than one packet it's also useful to record the seed packet number.

Standardise labelling

Don't label the packets arbitrarily: use a format which is followed on all seed packets. You could for instance record the accession number and underneath this put the batch reference, crop name and location. If you're using adhesive labels make sure the labels can't fall off the packets. If you're using a marker pen for labelling the packets, use an indelible ink which is not erased by seed packets rubbing against each other in the store. Also avoid labelling the packets near the top as the labels might get damaged when the packets are opened and re-sealed.

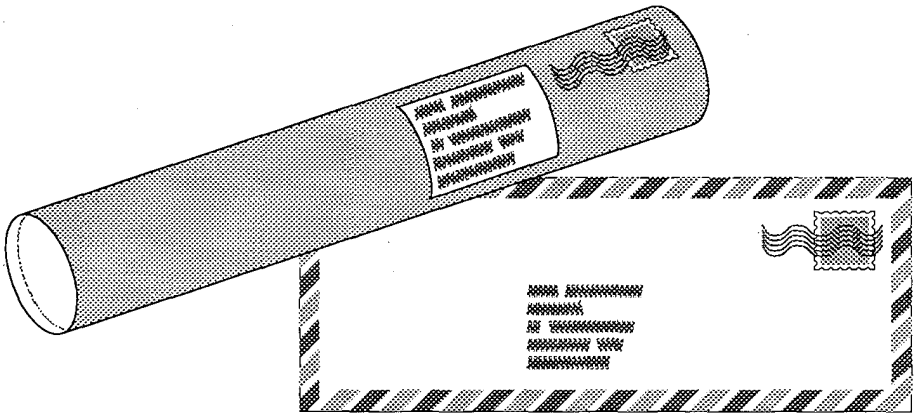
Organise the seed store sensibly

It can be hard work hunting round in a cold room or freezer for a seed packet, that's why each batch of an accession often has a seed store location address, usually in the form of a code. (e.g. "A/01/02/01": cold room "A", rack "01", shelf "02", box "01"). A major nightmare for documentation staff and other genebank staff alike is when-the seed store has to be re-arranged because of operational difficulties. Such re-organisations are sometimes necessary but in many cases can be avoided if careful thought is made how the seed store is organised in the first place e.g. should the seed store be arranged in strict accession number order or on a crop basis? How many racks; shelves and boxes are being used? In any case, the smallest container will be a box or drawer so it helps enormously if each drawer/box of seed packets is sorted in an accession number order.

Keep the seed store tidy!

Don't let seed packets fall out of trays or get lost. Remember, the documentation system is supposed to give information about the seed store: what germplasm it has and where. The documentation system won't tell you if a seed packet has fallen onto the floor (if it did, why are you documenting the fact that the seed packet is on the floor).

2.6 Seed distribution



2.6.7 Aim

To supply seeds of an accession together with relevant information in response to requests.

2.6.2 Outline

A request is received for a sample of a genebank accession. If there is sufficient quantity of seed in the seed store, a sample is removed and packaged for distribution. If there is insufficient seed in the store, the accession must first be multiplied before any seed is distributed. Passport and available scientific data are retrieved from the documentation system and supplied together with the sample.

The seed store inventory is updated to show how much seed remains for the particular batch reference of the accession. Details concerning the distribution are also recorded.

2.6.3 Descriptor list for seed distribution

Seed distribution has a strong documentation component so it's worth designing the seed distribution procedure so that documentation is an integral part and not an afterthought. If you're sending out samples on a regular basis, you'll make sure that the procedure is simple in de-

sign, well structured, efficient and easy to perform so that anyone can perform it and no mistakes can be made. To this end, you'll try to use standardised packages, standardised labels, standardised forms etc. in the procedure. These can save time in what is, after all, a routine operation.

The way that seed distribution data are documented is a good example of how genebank objectives and priorities and information needs directly affect the documentation system that is finally developed. For instance, if you're only sending out a few germplasm samples each year (and are unlikely to increase in the future) the chances are that you won't need a sophisticated and streamlined documentation system for seed distribution; perhaps a book or a file in a filing cabinet will suit your information requirements. At the other extreme, if you're sending out numerous seed samples every year and want to keep track of, say, where samples are going (and from which batch), how many accessions are sent out to different categories of scientists and so on, you'll want to keep a more complex documentation system and probably on a computer. The system might have similarities to management systems found in some larger retails shops or mail-order firms which require detailed sales analysis, stock control and invoicing/accounts features; however it's unlikely that you'll need all of these features. A descriptor list for seed distribution is given in table 7.

Table 7. Descriptor list for seed distribution

DESCRIPTOR	COMMENTS
Accession number	
Batch reference	
Date of supply	
Amount of seed sent	
Reference to recipient address	
Phytosanitary certificate number	
Export permit number	
Recipient's import permit number	
Mail registration number	

2.7 Seed monitoring

2.7.1 Aim

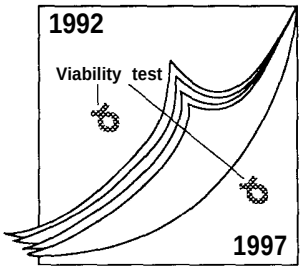
To establish whether seed amount and seed viability remain at acceptable levels.

2.7.2 Outline

Seed monitoring consists of two separate procedures:

1. To determine at pre-set intervals (e.g. every 5 years) the viability of accessions using a germination test. If viability is unacceptably low, regeneration is necessary.
2. To determine at regular intervals the amount of seed of each accession in the seed store. When seed amount is unacceptably low, the accessions should be multiplied.

2.7.3 General remarks



The sequence of events in seed monitoring is:

a. Determine accessions to test
b. Perform test
c. Act on result

Seed monitoring relies heavily on the documentation system to provide: information on which seed stocks are getting low, which accessions need viability testing and which accessions need multiplying/regenerating. If it wasn't for the documentation system providing good, up to date information, seed monitoring would be very difficult to carry out effectively.

2.8 Regeneration/multiplication

2.8.1 Aim

To replenish stocks of an accession by growing the accessions under suitable conditions.

2.8.2 Outline

Regeneration is performed when the viability of a batch falls below an acceptable level; this level is commonly 85%.

Multiplication, on the other hand is performed when the total weight or number of seeds of an accession falls below a pre-defined minimum amount.

The procedures developed for multiplying or regenerating a sample will depend on a variety of factors including the species, its breeding system and the heterogeneity of the accession. These are not covered here, but the procedures have the following common elements:

- Determine how many accessions can be handled

When the requirements for the different samples are taken into consideration along with the amount of land available, an approximate figure can be calculated for the number of accessions that can be regenerated or multiplied in the space available.

- Decide which accessions to regenerate or multiply

Priority is usually given to accessions in the following order:

1. Those which have a low viability
2. Those which are required urgently for supply or use within the genebank
3. Those which have a low seed number but high viability

Determine where to grow the material

For those accessions which are being regenerated, the passport data must be consulted to determine the appropriate conditions for regeneration.

- Prepare the sowing plan

Commonly, a map is drawn of the area to be used showing the allocation of accessions to particular plots.

- Prepare the seeds for sowing

The required quantity of seeds is removed from the store room and the seed distribution and inventory files adjusted accordingly.

- The samples are grown out for regeneration or multiplication
- Harvest and process the new seeds for entry into the genebank

Samples are harvested and then undergo the operations of seed cleaning, seed drying, seed viability testing, seed packing and storage as outlined earlier.

2.8.3 General remarks

Regeneration or multiplication is performed as a result of information generated in seed monitoring procedure. Passport and other data on the accessions are also required-in the planning of regeneration. Information is required (where appropriate) on the number of plants, the spacing in the plots, the breeding system of the crop, any isolation required

and the pollination method needed. Your genebank will probably already have these data recorded as standard methods so you may not have to concern yourself with recording these data for each accession.

If you have the possibility of regenerating accessions at more than one location you could consider listing the "preferred site of regeneration" in the inventory file. This will save you having to consult the passport file every time you have to decide where to regenerate an accession and make the planning process that much easier.

2.8.4 Descriptor list for regeneration/multiplication

Some or all of the descriptors in table 8 can be used for recording regeneration/multiplication data. As the methods vary according to species, there will also be a number of other species-specific descriptors which you will have to record. These are not given here.

Table 8. Descriptor list for regeneration/multiplication

DESCRIPTOR	COMMENTS
Accession number	
Batch reference	
Collection type	
Regeneration site	
Collaborator	
Plot reference	Field, row and plot number
Sowing date	
Transplanting date	
Sowing density	
Germination in the field (%)	
Number of plants established	
Days from sowing to flowering	
Breeding system	
Harvest date	
Cultural practices	

2.9 Characterisation/preliminary evaluation

2.9.1 Aim

To grow accessions under suitable conditions and to obtain data for particular descriptors.

2.9.2 Outline

The difference between characterisation and evaluation is as follows. Characterisation consists of recording those descriptors which are highly heritable, can be seen easily by eye and are expressed in all environments (e.g. flower colour, number of leaves). An example, the characterisation of the arrangement of storage roots on underground stems in sweet potato is illustrated in fig. 4. Evaluation consists of recording those characters which are susceptible to environmental differences (e.g. fruit yield, drought susceptibility). Therefore an accession can be evaluated on a number of different sites, perhaps giving significantly different results for several descriptors. On the other hand, an accession being characterised on several different sites is likely to give essentially the same results. You probably wouldn't characterise an accession more than once unless you were checking its integrity and wanted to know whether it still represented the original deposit.

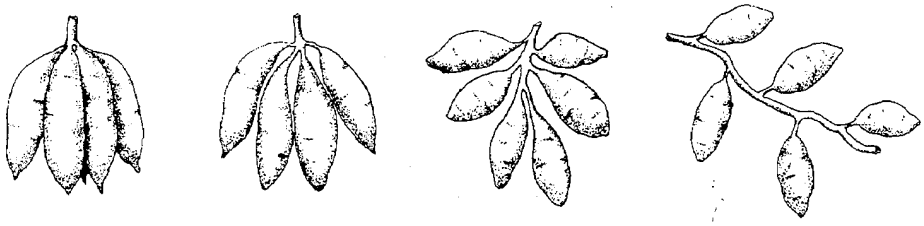


Fig. 4. Characterisation of the arrangement of storage roots on underground stems in sweet potato

In practice, when an accession is being characterised, a limited number of evaluation characters are recorded at the same time (so called "preliminary evaluation" characters). These are generally fairly straightforward to record and are of general interest to users of the particular crop.

A number of different procedures are used for characterisation and preliminary evaluation depending on the species. The procedures vary in operation and how labour intensive they are. The characterisation or preliminary evaluation trials would be planned in a similar way to regeneration or multiplication procedures and are likely to have the following steps:

- Determine how many accessions can be handled
- Decide which accessions to characterise or evaluate
- Determine where to grow the material
- Prepare the planting plan
- Prepare the seeds for planting
- Perform characterisation or preliminary evaluation
- Record data in documentation system

Although characterisation can be performed at the same time as regeneration or multiplication, evaluation cannot; for instance, you wouldn't assess the pest or disease tolerance of an accession when regenerating it.

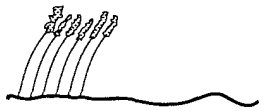
2.9.3 Descriptor list for characterisation/preliminary evaluation

Most of the descriptors for characterisation and preliminary evaluation are species-specific. However, there are some general descriptors concerning the conditions of the trial which are given in table 9.

Table 9. Descriptor list for characterisation/preliminary evaluation

DESCRIPTOR	COMMENTS
Accession number	
Batch reference	
Site (research station)	
Latitude	
Longitude	
Altitude	
Evaluator(s)	
Sowing date	
Harvest date	
Evaluationenvironment	
Soil type	
Soil pH	
Monthly rainfall	
Cultural practices	

3 Field collections



The management and documentation of field collections shows great similarities with that of seed collections. For this reason we wont go into the same detail here as many of the concepts have already been discussed. One big difference is that in routine management procedures there is often a need to document information on the individual plants that make up an accession.

The common procedures undertaken at field collections are visualised in fig. 5.

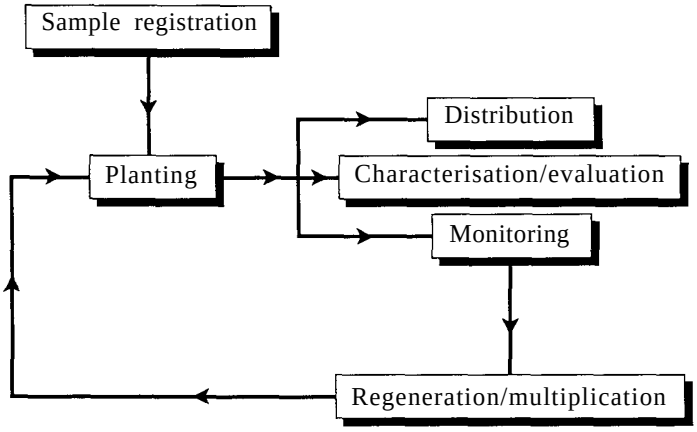


Fig. 5. Common procedures undertaken at field collections

3.1 Sample registration

The same considerations apply here as for the registration of seed samples. If there is a future need to document information on the individual plants, each plant in the accession should be given a unique plant identification number. Quarantine regulations are also strictly observed since field collections are more susceptible to pests and diseases than seed collections.

3.2 Sample planting

Healthy accessions are planted according to a pre-arranged plan in the field, glasshouse or orchard. The field inventory will therefore be a record of the precise location of the accessions (and the individual plants). The descriptors might include some or all those listed in table 10.

Table 10. Descriptor list for field collection inventory file

DESCRIPTOR	COMMENTS
Accession number	
Batch reference	
Plant identification number	
Plot reference	Field, row and plot number
Date of next regeneration	

3.3 Sample monitoring

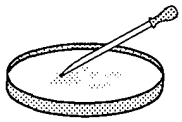
Accessions are checked periodically for disease or infestation with pests. Any diseases or pests are dealt with according to standard procedures. Information about any treatments can be documented for management purposes.

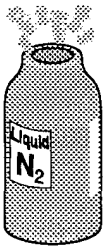
3.4 Other activities

- Sample distribution
- Characterisation/evaluation
- Regeneration/multiplication

Most of the considerations for seed collections also apply to field collections with regard to these activities. There will often be a need to document the plant identification number routinely in these procedures.

4 *In vitro* collections





Germplasm is maintained *in vitro* as plant tissue ranging from protoplast and cell suspensions to callus cultures, meristems/shoot-tips and embryos. A variety of different methods are used for maintenance, including culture methods on solid or liquid medium and storage at very low temperatures *(cryopreservation).* As in the case of field collections, the management and documentation of *in vitro* collections show similarities to that of seed collections but there are some important additional considerations, principally concerned with the way the germplasm is maintained.

The procedures commonly undertaken by in vitro active collections are shown in fig 6.

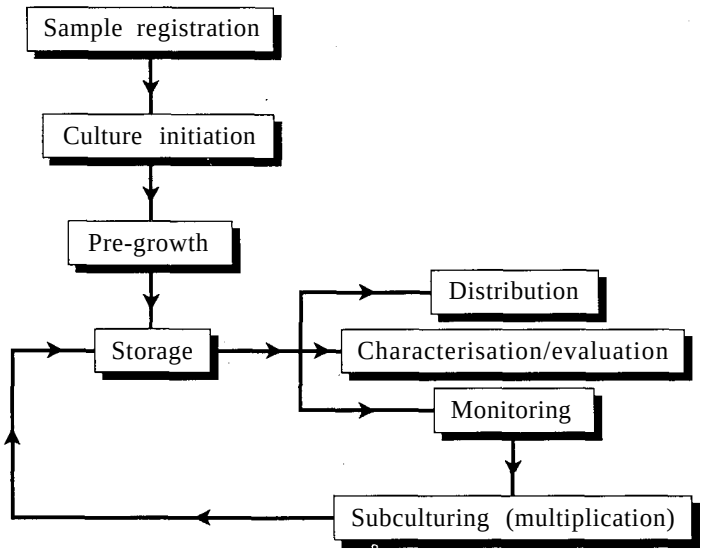


Fig. 6. Procedures commonly undertaken in in *vitro* active collections

Samples may enter the *in vitro* genebank from other *in vitro* collections or as cultures from field-grown material, newly collected vegetative samples, or seeds. As we saw with field collections, vegetatively propagated material frequently presents more serious plant health and quarantine problems than seeds since it tends to accumulate those pathogens which are otherwise removed at the seed production stage. Therefore, the initiation of cultures for *in vitro* conservation usually involves disease indexing and eradication.

In vitro storage may be short to medium-term under normal culture conditions or conditions that reduce the growth rate ("slow growth storage"). For long-term storage, cryopreservation in liquid nitrogen is used. The former is a cyclical process with subculturing every, say, 1 or 2 years (see fig. 6), whereas the latter is a one-step storage method for which the only recurrent activities (other than addition or distribution of samples) are periodic monitoring for viability and stability. *In vitro* cultures are often duplicated in parallel under different conditions, e.g. slow growth and cryopreservation, or slow growth and in the field. They may also have related seed collections.

Cultures in normal or slow growth require monitoring periodically to note features such as microbial contamination, callus formation, loss of leaves, bleaching of leaves and browning of the culture on the medium. The timing of subculturing and the number of replicate cultures

produced is influenced by data obtained during the monitoring process. For example, cultures showing signs of excessive browning require early subculturing; those affected by contamination should be discarded and replaced by multiplication of clonally identical replicate cultures.

The scope for characterisation of *in vitro* is extremely limited except for biochemical and molecular techniques, e.g. isozyme studies and DNA analysis using *Restriction Fragment Length Polymorphism (RFLP)* or *Random Amplified Polymorphic DNA (RAPD).* This is because when vegetative material is introduced into culture, it looses many of its distinguishing morphological features that it had as an independently growing plant. Therefore, full characterisation is usually carried out prior to culture initiation or using field-grown, clonal replicates of cultured material. Also, because of the risks of instability, genetic stability is commonly monitored by whatever means are available. This may involve isozyme analysis, molecular screening (RFLP or RAPD) or periodic transfer to the field for evaluation.

The batch reference of an accession is of particular significance in *in vitro* collections and is always recorded. Often the batch reference represents the *pedigree* of the sample and commonly uses the notation illustrated in fig. 7.

Genetic instability

Continued genetic stability of the accessions maintained is a matter of concern for all genebanks; with *in vitro* collections particular care needs to be taken to guard against instability. Unorganised cultures such as cell suspensions are more prone to genetic instability (or *somaclonal variation)* than organised systems such as meristems and embryos. The latter are therefore favoured for genetic conservation but callus formation could still occur in storage, increasing the risk of instability. Also, shoot cultures of some crops such as banana and plantain are inherently unstable, often amplifying the somatic mutations that can occur in the field.

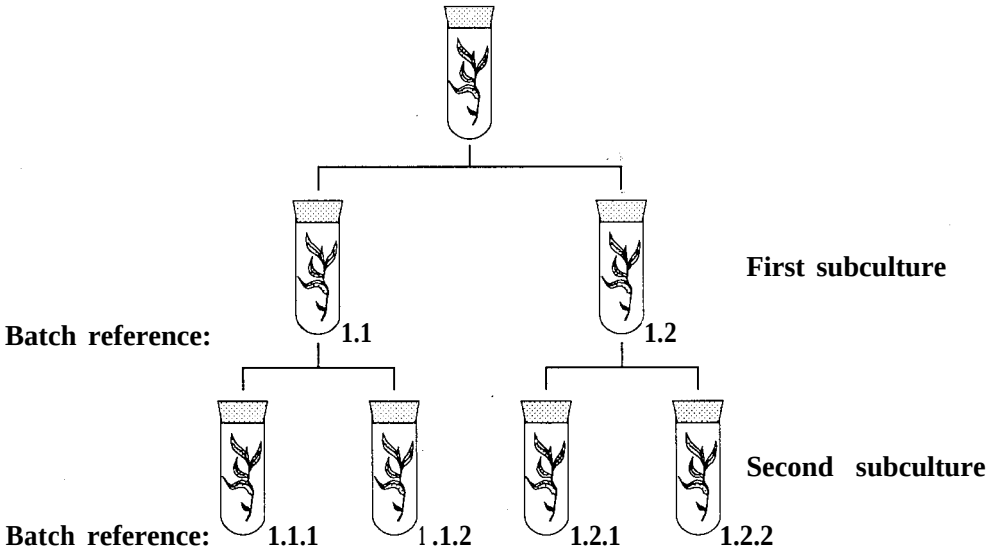


Fig. 7. An example of the batch reference notation in *in vitro* collections

The notation is rather like the numbering system used in this guidebook to distinguish the different sub-sections of each chapter. Alternatively the batch reference can be a reference to the subculturing regime used (either systematic or random).

There are many aspects of this *in vitro* conservation process that require detailed analysis when constructing the documentation system. Table 11 summarises descriptors from a number of different procedures which are important for the management of *in vitro* collections. You should perform detailed analysis of each procedure to produce your own descriptor lists. Don't forget that you can document the passport data in the same way as for seed collections. Also, precise details of the various techniques used (culture medium and conditions, disease indexing, monitoring, etc.) can be kept in separate files and a reference made to these from the inventory files.

Table 11. Descriptors commonly used for management of *in vitro* collections

DESCRIPTOR	COMMENTS
Accession number	
Batch reference	
Relationship to other conserved material	Note on whether sample was derived from other conserved material
Date of culture initiation	
Disease indexing procedures	Reference to which procedures were carried out
Disease indexing results	
Disease eradication procedures	Reference to which procedures were carried out
Number of replicates maintained in culture	
Culture medium used for initiation	
Culture medium used for maintenance	
Storage conditions	Temperature, day length, etc.
Normal subculturing frequency	
Normal monitoring frequency	
Last subculturing date	
Next subculturing date	
Subculturing regime	Random or systematic
Storage location(s)	
Location of safety duplicates	

5 Questionnaire/analysis

The following questionnaire is designed to help you analyse the procedures carried out in your genebank; to identify data generated in the course of each procedure, and to identify meaningful sets of descriptors for each procedure performed. It is essential to spend time completing this questionnaire, for each procedure in turn, as it will be used as the basis of your new documentation system.

We looked at the different procedures commonly carried out in genebanks in the previous sections of this chapter and identified different descriptors commonly used for recording data on specific procedures. These will be a useful reference for completing your questionnaires.

The procedures carried out at your genebank may differ from the procedures we have studied. There may be more procedures you wish to document, or certain procedures which need to be documented in a different way.

The first stage of this analysis is to identify all procedures performed at your genebank. You may have identified these already. If not, the following list details all the procedures examined earlier, which will provide a useful starting point.

Sample registration
Seed cleaning
Seed viability testing (germination test)
Seed packing and storage
Seed distribution
Seed monitoring
Regeneration/multiplication
Characterisation/evaluation

You should construct a flow chart showing how the different procedures are related. Refer back to Chapter 3 and the earlier part of this chapter if you are unsure how to do this.

For each procedure you have identified, you should start a separate analysis form. The form given on page 85 is designed specifically for this purpose. Photocopy the form so you have one for each procedure identified. At this stage, complete only the preliminary details on each form: the procedure you are analysing, the name of your genebank, the location of your genebank, your name and today's date.

The next step is to decide which descriptors need to be recorded in your documentation system for each procedure identified. You will find it useful if you construct a flow chart for each procedure. Look at the data produced in each individual procedure and consider carefully which

descriptors need to be recorded. Again, you may have identified these already. If not, look back at the previous sections of this chapter. The descriptors listed for each procedure discussed are a useful starting point.

Work through the procedures one at a time. For each procedure, as you identify a given descriptor, this should be added to the **"DESCRIPTOR"** column on the form. If you know of any descriptor states that are currently used when recording a particular descriptor, it is useful to record these now in the **"DESCRIPTOR STATES"** column. Any coding systems used should be indicated in the **"DESCRIPTOR STATES"** column, along with the descriptor state, and can be explained further in the **"COMMENTS"** column. If, however, you are unsure of descriptor states or coding systems used, or no formal systems currently exist, do not spend a lot of time working on them now, as they are studied in depth in the following chapter. They should be considered after Chapter 5 has been completed. Fig. 8 illustrates the different columns that must be completed.

DESCRIPTOR	DESCRIPTOR STATES	COMMENTS
ACCESSION NUMBER	GENEBANK NUMBER	CODING SYSTEM: OMPØØØØ – OMP9999

Fig. 8. Form for recording the descriptor list for a given procedure

The comments column can also be used for any miscellaneous data you might want to record about each descriptor, and a larger comments section is provided at the bottom of each form for you to record any miscellaneous data concerning the procedure as a whole. An example of a completed form is given in appendix 1.

If you are still unsure of where to start analysing your genebank procedures, table 12 at the end of this chapter should help you. It lists all the procedures that we have discussed in this chapter, suggesting meaningful descriptors that could be recorded in each case.

You are advised to consult table 12 only after you have given careful thought to the procedures in your own genebank. It is vital that each form is based on how procedures in your genebank operate. All genebanks operate in different ways, so it is impossible to produce a standard model.

GENEBANK PROCEDURE ANALYSIS

NAME OF GENEBANK:

LOCATION OF GENEBANK:

YOUR NAME:

DATE:

PROCEDURE:

TYPE OF PROCEDURE (OPERATIONAL/SCIENTIFIC):

RELATED PROCEDURES:

SHEET NUMBER:

DESCRIPTOR	DESCRIPTOR STATES	COMMENTS

COMMENTS:

GENEBANK PROCEDURE ANALYSIS

CONTINUATION SHEET

PROCEDURE:

SHEET NUMBER:

DESCRIPTOR	DESCRIPTOR STATES	COMMENTS

COMMENTS:

Table 12. **Meaningful descriptors that can be recorded for different procedures. Further crop specific descriptors can be found in descriptor lists published by IBPGR**

DESCRIPTOR	ACCESSION	COLLECTING	CLEANING	DRYING	VIABILITY	INVENTORY	DISTRIBUTION	REGENERATION
Accession number	✔	✔	✔	✔	✔	✔	✔	✔
Batch reference			✔	✔	✔	✔	✔	✔
Collection type					✔	✔		✔
Scientific name	✔	✔						
Pedigree/cultivar name	✔							
Donor name	✔							
Donor identification number	✔							
Other numbers associated with the accession	✔							
Acquisition date	✔							
Date of last regeneration or multiplication	✔							
Collecting number		✔						
Collecting organisation		✔						
Expedition identifier		✔						
Collecting date		✔						
Country of collecting		✔						
Province/state		✔						
Location of collecting site		✔						
Latitude of collecting site		✔						
Altitude of collecting site		✔						
Origin (if different from collecting site)		✔						
Collecting source		✔						
Status of sample		✔						
Type of sample		✔						
Local/vernacular name		✔						
No. of plants sampled		✔						
Land use		✔						
Vegetation type		✔						

Table 12 (cont.). Meaningful descriptors that can be recorded for different procedures. Further crop specific descriptors can be found in descriptor lists published by IBPGR

DESCRIPTOR	ACCESSION	COLLECTING	CLEANING	DRYING	VIABILITY	INVENTORY	DISTRIBUTION	REGENERATION
Genetic erosion		✔						
Plant use		✔						
Cultural practices		✔						✔
Photograph identification number(s)		✔						
Herbarium identification number		✔						
Topography		✔						
Slope		✔						
Aspect		✔						
Annual rainfall		✔						
Rainfall seasonality		✔						
Soil type		✔						
Soil pH		✔						
Soil texture		✔						
Date seed cleaned			✔					
Reference to cleaning method			✔					
Estimate of total seed			✔					
Proportion of empty seed			✔					
Treatment of seed			✔					
Operator (cleaning)			✔					
Reference to drying method				✔				
Final moisture content determination (%)				✔				
Date of moisture content determination				✔				
Total dry weight of seeds				✔				
1000 seed weight				✔		✔		
100 seed weight (for large seeds)				✔		✔		
Reference to viability method					✔			

Table 12 (cont.). Meaningful descriptors that can be recorded for different procedures. Further crop specific descriptors can be found in descriptor lists published by IBPGR

DESCRIPTOR	ACCESSION	COLLECTING	CLEANING	DRYING	VIABILITY	INVENTORY	DISTRIBUTION	REGENERATION
Date of viability test					✔			
Viability (%)					✔			
Operator (viability)					✔			
Location in seed store						✔		
Total amount of seed						✔		
Minimum amount of seed allowed						✔		
Number of packets/containers						✔		
Date of supply							✔	
Amount of seed sent							✔	
Reference to recipient address							✔	
Phytosanitary certificate number							✔	
Export permit number							✔	
Recipient's import permit number							✔	
Mail registration number							✔	
Regeneration site								✔
Collaborator								✔
Plot reference								✔
Sowing date								✔
Transplanting date								✔
Sowing density								✔
Germination in the field (%)								✔
Number of plants established								✔
Days from sowing to flowering								✔
Breeding system								✔
Harvest date								✔

Chapter 5

Data recording

Chapter 5 will look in more detail at how data can be recorded. When you have finished this chapter, you will be able to:

- Review the advantages and disadvantages of the different approaches to recording data
- Explain why descriptor lists are used
- Illustrate the different steps in developing descriptor lists
- List the conventions used when scoring descriptors
- Describe ways of handling heterogenous data

1 Approaches to recording data

In Chapter 4 we analysed in some detail the common genebank procedures and identified sets of descriptors that could be used in a genebank documentation system. However, we haven't looked in much detail about how the data are recorded or the factors which can affect the accuracy and precision of the data. Some or all of the steps listed in sections 1.1 to 1.6 are followed when data are collected.

1.1 Observe or measure trait

It's important that the person making the observation is familiar with the trait or characteristic that is being studied otherwise the data obtained will be of doubtful quality. For example, if observations are being made on the colour of seedling hypocotyls and the observer does not know what a seedling hypocotyl is, it's highly unlikely that the data obtained will be reliable. However, familiarity with the trait by itself is not enough. The observer also has to know the most accurate way to make the observation, usually by employing standard methods. If the observer is careless in making the observation or the equipment being used is faulty, the data obtained will be unreliable. For instance, if a germination test is carried out using the wrong method, any viability data will be questionable. Similarly, if a hand-held moisture content meter is known to be faulty, can you trust any measurements made for seed moisture content?

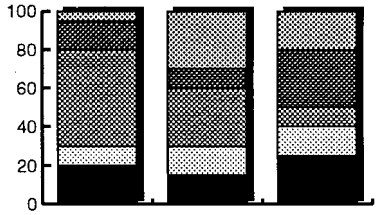
1.2 Record observation or measurement

Raw data are commonly recorded on paper, often using standardised forms. In many applications it is also possible to use computerized data loggers which greatly facilitate this often laborious process. For instance, observations in a field trial can be recorded directly using a hand held data logger. Also bar codes, which are a familiar feature on many commercial packages, are increasingly being used for seed store inventories as they simplify management of the collections. Whether using manual or computerised methods for recording raw data, it is essential that the data are recorded accurately and with the desired degree of precision.

1.3 Transcribe into computer or manual format

Data which have been recorded can be transcribed into a different format (e.g. from field notebooks to manual forms) or into different media (e.g. from field notebooks to a computerised format). Data transcription is a common source of errors if it is not performed with great care. For this reason, the number of different transcriptions is usually kept to a minimum, particularly between different manual formats.

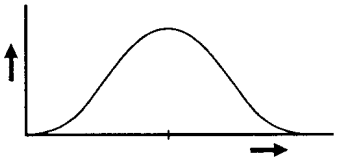
1.4 Analyse raw data



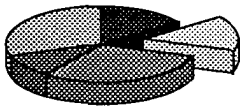
Data need to be analysed or transformed to produce information. A wide variety of analyses is possible ranging from simple arithmetic calculations to complex statistical analyses. The analysis performed is entirely dependent on the traits being studied and the information required from the analysis.

1.5 Score descriptor using results of analysis

Many descriptors can be scored directly from the raw data and require no analysis. Examples of these descriptors include seed weight, collection source and date of next viability test. However, the majority of descriptors concerning scientific data and many management descriptors require some sort of analysis of the raw data before scoring the descriptors. For instance, the majority of characterisation descriptors are based on observations or measurements of, say, 20 plants; these data need to be analysed before the descriptor can be scored. Calculations which are commonly recorded include the mean, standard deviation, median and mode.



1.6 Output data into a more usable format



Once the descriptors have been scored, the data can be output into other formats depending on the application. In computerised systems, for example, data can be output to spreadsheet or statistical software for further analysis. For the purpose of information dissemination, data are commonly organised to produce specific reports which might contain tables, graphs; pie-charts and other graphical elements.

2 Advantages and disadvantages of different approaches

There are a number of approaches to recording data. The approach which is chosen will affect:

- The ease with which data can be recorded
- The ease with which recorded data can be updated/modified
- The ease of flexible information retrieval
- The potential for data analysis/transformation

The different approaches are summarised in tables 1 and 2.

Table 1. Different approaches to recording quantitative data

APPROACH	EXAMPLE	NOTES
Continuous scale	Plant height: 0.9m	Descriptor is scored using standard (SI) units (e.g. metres, grams)
Ordinal scale	Plant height (9 descriptors): 1 = very short (<0.5m) 2 = very short to short (>0.5-0.75m) 3 = short (>0.75-1.0m) 4 = short to intermediate (>1.0-1.25m) 5 = intermediate (>1.25-1.5m) 6 = intermediate to tall (>1.5-1.75m) 7 = tall (>1.75-2.0m) 8 = tall to very tall (>2.0-2.25m) 9 = very tall (>2.25m)	Descriptor is scored using a series of pre-defined descriptor states. In this example, a plant height of 0.9m is scored as either “short” or "3"
Binary scale	Plant height (9 descriptors): Presence of very short (<0.5m) plants Presence of very short to short (>0.5-0.754) plants Presence of short (>0.75-1.0m) plants Presence of short to intermediate (>1.0-1.25m) plants Presence of intermediate (>1.25-1.5m) plants Presence of intermediate to tall (>1.5-1.75m) plants Presence of tall (>1.75-2.0m) plants Presence of tall to very tall (>2.0-2.25m) plants Presence of very tall (>2.25m) plants	Each descriptor has two descriptor states: + (present) and 0 (absent). In this example, the single descriptor for plant height has been replaced with 9 separate descriptors for the individual ranges, each scored as + (present) or 0 (absent)

Table 2. Different approaches to recording qualitative data

APPROACH	EXAMPLE	NOTES
Nominal scale	Flower colour (8 descriptors): 1 = white 2 = cream 3 = yellow 4 = orange 5 = green 6 = dark green 7 = red 8 = dark red	Descriptor is scored using a series of pre-defined descriptor states. In this example, a plant producing orange flowers is scored as "orange" or "4"
Binary scale	Flower colour (8 descriptors): Presence of white flowers Presence of cream flowers Presence of yellow flowers Presence of orange flowers Presence of green flowers Presence of dark green flowers Presence of red flowers Presence of dark red flowers	Each descriptor has two descriptor states: + (present) and 0 (absent). In this example, the single descriptor for flower colour has been replaced with 9 separate descriptors for the individual ranges, each scored as + (present) or 0 (absent)

2.1 Continuous scales

Certain descriptors *have* to be scored on a continuous scale. A good example is seed weight – you need to know precisely how much seed is in the cold store. However, before using a continuous scale a decision has to be made about the precision when scoring the descriptor – how many decimal places are required: 1, 2, 3, 4, none? Should you record a seed viability as 87.342% or just 87%? Should you record a seed weight as 256.13g or as 256g? Any decision you take here can affect the ease of data entry and modification and the usefulness of any information retrieved.

Quantitative data have a greater potential for statistical analysis on a continuous scale than on an ordinal scale. Therefore, if statistical analysis is a priority, a continuous scale should be used when scoring the descriptors. Transformation from a continuous scale to an ordinal scale is possible but with an accompanying loss of precision. For instance, a plant height of "0.9m" can be transformed to an ordinal scale as "small" but it is not possible to do this in the other direction. Therefore it's best to record the raw quantitative data on a continuous scale, this way the information potential of the data is maximised.

2.2 Ordinal scales

As with continuous scales, scoring descriptors using an ordinal scale is usually quite straightforward. Future updating or modification of these data in a documentation system is easy since there are only a limited number of descriptor states. An ordinal scale is illustrated in fig. 1.

The potential for data analysis and transformation is not so great as with continuous scale data, However, certain types of information retrieval are facilitated since the data have, to a certain extent, been interpreted and classified. You'll easily be able to search directly for "disease resistant plants", "high yielding varieties". You'll be able to assess more directly how similar/dissimilar individual genebank accessions are. However, you'll only be able to do this if the descriptor states are meaningfully defined for a particular crop. This is an important point and we'll be returning to it later in section 3.2.3.

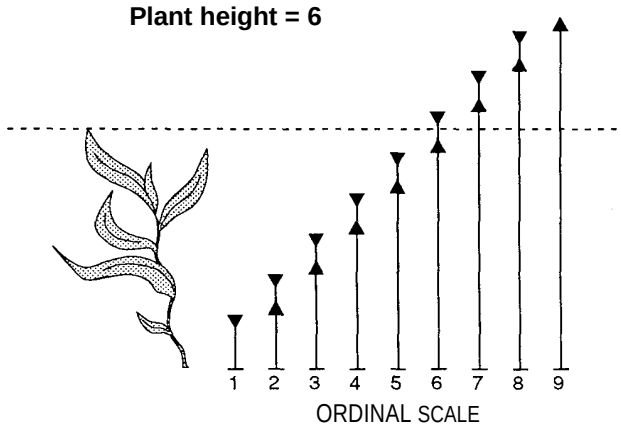


Fig. 1. Scoring plant height using an ordinal scale

2.3 Use of numeric codes

The use of numeric codes greatly facilitates the simple and accurate scoring of descriptors. For instance, it would be time consuming if you had to type or write "high yielding variety" for a whole series of accessions: it would take up a lot of space on the forms you are using or in the computer database. Mistakes might creep in: you might misspell a word or later on find someone else's handwriting difficult to read. These mistakes would make information retrieval more difficult. It's a lot easier to type or write "2" than to write "high yielding variety". Data updating and modification are easier as well – and it's much easier to search for "2"!

The main problem with numeric codes arises when you are handling more than one crop. Each crop has its own descriptor list and these are not necessarily related to each other, particularly with regard nominal scales such as "uses of accession" or "collection site". When you consult the documentation system for information, what do all those "2"s mean? (And what do all those other numbers mean – 1, 3, 5, 2, 6, 4, 8, 9 ?). You need to go back to all those pre-defined lists to work out what they mean. Information retrieval can become very slow, particularly in manual systems. In computerized systems, the computer software sometimes has facilities for automatic conversion of the numeric codes to the corresponding descriptor state so that you can search for (or display) either "5" or "intermediate" and so on for the other descriptor states.

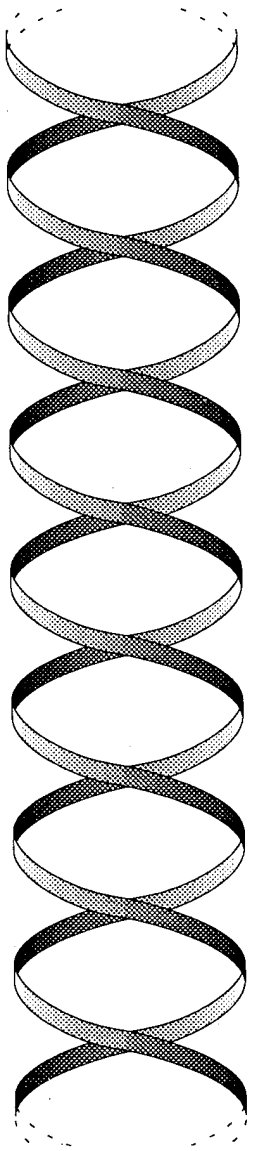
2.4 Use of letter codes

You should ***never*** consider using short-hand codes instead of numeric codes e.g. "vs" for very short, "grn" for green, "mkt" for market, "I" for irrigated etc. The problem with short-hand codes is that, while they might be obvious to you, they are not always obvious to other people, especially you use short-hand codes for many descriptors. Worse still, different people using the documentation system will use different codes leading to confusion e.g. "vs", "v.sh", "v.sht" for very short. The lack of uniformity will have serious consequences for information retrieval and information exchange. Just think of the problems of exchanging data with other national genebanks, all using short-hand codes: for the English "very short" might be "vs", the Spanish "muy corto" – "mc" and the French "très courte" – "tc". What a confusion! On the other hand, with a numeric coding these might all be "1". Even if you use short-hand codes in the original lab-books, the documentation system itself should contain only numeric codes (or their translated equivalent).

2.4.1 Country codes

There are occasions where non-numeric, letter codes can be used but these are comparatively rare unless you are working with genetic data. An example is the 3 letter codes for country names recommended by the Statistical Office of the United Nations. These abbreviations are internationally recognised and widely used and can be found in appendix 2.

2.4.2 Gene symbols



With genetic data, symbols are widely used to represent particular traits. These symbols can often be quite long and confusing and the conventions used can vary depending on the biological system under study (e.g. bacteria, viruses, fungi, plants etc.). Nonetheless, a number of conventions are commonly followed in the documentation of genetic data in plant genetic resources. Some of these are now given:

1. A common basic symbol is given for a particular trait. When the trait can be caused by more than one gene, the different genes are numbered sequentially using Arabic numerals. For instance; in maize the symbol *dek* stands for the trait "defective kernel"; more than 30 genes can be responsible for this trait and these are numbered in sequence *dek1*, *dek2*, *dek3* and so on. Notice that the symbols are written in italics; alternatively they can be underlined.
2. Where there is no ambiguity, the basic symbol of a dominant gene should begin with a capital letter and those of a recessive gene with a small letter. This information is of particular significance when performing genetic crosses. So in maize, the trait "defective kernel" is recessive and written as *dek* whereas the trait "wrinkled kernel" is dominant and written *Wrk*.
3. Different alleles of a particular gene are designated by a lower case character following the gene designation. For instance, in rye the designations *Pm1a* and *Pm1b* represent different alleles of the same gene *Pm1* which is responsible for the trait "resistance to powdery mildew". Where the alleles have been detected by isozyme analysis, it is common to separate the allele designation from the gene symbol with a dash. Therefore, in soybean alleles of the gene *Sod2* (for superoxide dismutase) are written *Sod2-a, Sod2-b*.

There are other conventions used which are more specific to the species under study so they will not be given here. It should be pointed out that many of the symbols used are only meaningful within the species under study. For example, genes are designated as they are identified so although the symbol *Pgm* might be used-for the enzyme phosphoglucose mutase in a number of species, the numbering of the genes is particular to the species. Therefore, care should always be taken when documenting genetic data from different species. Use accepted and well established conventions – don't create your own! A considerable number of symbols are already used with genetic data and you don't want to add to the confusion by creating more!

2.5 Binary scale

In theory, descriptors scored on a binary scale are by far the easiest to score, update and modify in a documentation system. You don't need to consult any pre-defined lists. It's just a matter of answering "present"/"absent" ("yes"/"no"). Information retrieval is enhanced as the descriptor states are so rigorously classified and easy to compare; data exchange is also much easier.

You might run into practical problems when scoring descriptors if there are many descriptors scored using a binary scale. As we saw from tables 1 and 2, descriptors scored using an ordinal or nominal scale can be expressed on a binary scale with each descriptor state being a separate binary observation. So, 20 descriptors scored on an ordinal scale with 9 separate descriptor states would correspond to 180 descriptors scored on a binary scale. That's 160 more entries to make on computer or on manual forms. Quite an increase in the workload for scoring descriptors! It would also be more difficult to retrieve information. For instance, it would be more difficult to answer general questions like "what colour flowers does this accession produce?" Instead you'd have to ask "does this accession produce red flowers?" "Does this accession produce blue flowers?" And so on.

In some cases, using a binary scale would waste a lot of time and effort. If the crop regularly produces a variety of different flower colours in the same accession, then the different descriptor states can be recorded efficiently on a binary scale by indicating the presence or absence of each colour. On the other hand, you wouldn't use a binary scale for nominal scale descriptors such as "collection site". The fact that an accession was collected in a "village market" would necessarily exclude its collection from a "research institute". If there were 9 separate collection sources, 8 of these would *always* be recorded as negative.

The use of binary scales is limited

Statistical analyses or data transformations on binary scale data are more limited than those for ordinal/continuous scale data. For these reasons, and the reasons given above, the use of data scored on binary scales is limited.

3 Descriptor lists

3.1 The need for descriptor lists

Why do you need to use descriptor lists? For instance, with the descriptor "collection site", why not just score the descriptor as it is: after all, you don't need a list to tell you that you collected a sample of maize in a "village market", do you?

Descriptor lists for scoring descriptors will considerably simplify **ALL** operations

The answer to this question is simple. The use of well-defined, tested and rigorously implemented descriptor lists for scoring descriptors will considerably simplify *all* operations concerned with data recording, updating/modifying, information retrieval, and exchange, and data analysis and transformation. In other words, you must use an existing and widely accepted descriptor list!

When data are recorded, they have to be classified and interpreted; with a pre-defined list of descriptors and descriptor states to consult, it saves a considerable amount of time. The use of lists ensures uniformity within the genebank (and between genebanks) and reduces error. People using the same lists at different genebanks will be able to exchange data readily and interpret the data with few, if any, problems.

The need for descriptor lists is highlighted in table 3. Two genebanks want to exchange data. However, both genebanks are using different scales for the scoring descriptors. Data exchange is clearly a problem!

Table 3. Potential difficulties in data exchange between two genebanks using different descriptor definitions, scales or codes for scoring descriptors

DESCRIPTOR	GENEBANK 1	GENEBANK 2	COMMENTS
Collection source	2 = farm land	2 = backyard	Different codes used for the nominal scale
Mature leaf size	9.5 cm	5	Different descriptor definitions as a result of the use of different types of scale: continuous and ordinal
Monthly rainfall	17 mm	1.7cm	Different descriptor definitions as a result of the use of different units.
Photograph taken	+	Y	Different codes used for binary scale

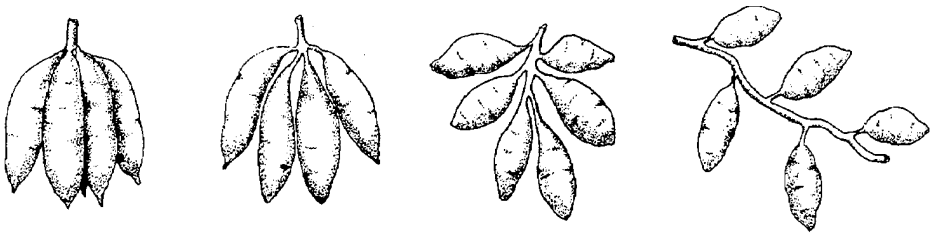
3.2 Conventions used in descriptor lists

Consider the following descriptor used for sweet potato:

6.1.1 Storage root formation

Arrangement of the storage roots on the underground stems.

1 Closed cluster
3 Open cluster
5 Dispersed
7 Very dispersed



1 Closed cluster 3 Open cluster 5 Dispersed 7 Very dispersed

Fig. 2. Storage root formation in sweet potato

In this example and many others you will come across in the literature, you will notice that the descriptor includes:

- A label **(6.1.1)** and a name (Storage root formation) – which clearly identifies the trait
- Descriptor definition – to assist in making the observation and/or subsequent scoring of the descriptor
- A list of descriptor states (where appropriate)
- Any units used (not relevant in this example)

Other conventions which are routinely used in plant genetic resources documentation are as follows.

3.2.1 Ordinal scales

- Descriptors are scored on a scale of 1-9

The great majority of descriptors scored using an ordinal scale can be scored on a scale of 1-9 where the descriptor states can be interpreted as follows:

1 = very low
2 = very low to low

3=low
4=low to intermediate
5=intermediate
6=intermediate to high
7=high
8=high to very high
9=very high

The use of this convention helps considerably in the recording, modifying, retrieval and interpretation of data from a wide number of descriptors. You will know that "5" is always "intermediate", "9" is always "very high" etc. An example of the use of this scale is stress susceptibility e.g. susceptibility to damage from low temperatures or high temperatures, soil acidity and susceptibility to damage by pathogens.

You will notice in published descriptor lists that often only a selection of the states is listed, e.g. 3, 5, 7. Where this occurs, the full range of the scale (1-9) is available by extension of the codes given or by interpolation between them.

▶ Use of other characters is restricted

There will be cases where it isn't possible to score a descriptor on a scale of 1-9 in which case the conventions listed in table 4 are commonly used.

Table 4. Use of other characters when scoring descriptors

SCORE	COMMENTS
0	No expression of descriptor. e.g. 0 is scored for seed colour when no seed are produced
+	Used for the ungraded prescence of a descriptor. e.g. + is scored for "seed colour" where seed are produced but no data are available on the seed colour produced
X	Where the descriptor is not uniform throughout the accession and so a number of different descriptor states are observed. This indicates that the accession is heterogeneous – either because the accession is a mixed collection or because genetic segregation with regard to the descriptor has taken place

It's important that the code zero (0) is reserved for cases where there is no expression of character, rather than the situation where the real answer is "don't know". Odd though it might sound, it's actually quite important to be aware of what you "don't know" about an accession or a group of accessions-you might want to score these descriptors at a later date. "Don't know" as a descriptor state is commonly indicated by recording a blank.

The use of the character "X" is not the only way of indicating heterogeneity within an accession – we'll look at other ways in section 4.

3.2.2 Continuous scales

It is important that any continuous scale is based on the International System of Units (SI) and that all units are used consistently. For instance, plant height for a particular species should be measured in either metres or centimetres, *not* sometimes metres, sometimes centimetres – and certainly never in inches!

3.2.3 Nominal scales

When descriptors are scored on a nominal scale using numbers rather than text (e.g. "2" instead of "farm land"), the use of 0, + and X should be restricted in the same way as-for ordinal scales. So, if "0" is scored, this will always mean "descriptor not observed". It's also helpful when the descriptor states are ordered in a logical order where appropriate. A good example is for the descriptor "Secondary flesh colour" in sweet potato:

0 Absent
1 White
2 Cream
3 Yellow
4 Orange
5 Pink
6 Red
7 Purple-red
8 Purple
9 Dark purple

3.3 Defining ordinal scales

Sometimes you will need to define the descriptor states for the descriptors you want to use when there are no descriptor lists published for the crops you are using or where the published lists are inadequate for your needs.

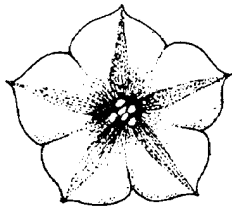
Statistical analyses and data transformations with ordinal scale data are not as powerful as the analyses with continuous scale data. This is because there is a loss of precision in the classification process. In order to maximise the possibilities for meaningful analyses (and therefore meaningful information retrieval), you should pay careful attention to how you construct the ordinal scales so that the inherent variation is not lost altogether. So, how do you determine the scale of 1-9 so that 1=very low and 9=very high for a particular descriptor? There are three things you could do:

- *Consultation* – ask the experts
- *Review the literature*
- Use *a continuous scale* in the first instance, subsequently analyse the data to find out the nature of the distribution

If the data are found to be normally distributed, you could construct an ordinal scale based on the mean and standard deviation. However, be careful when you construct this scale because some traits (such as yield and stress susceptibility) are sensitive to environmental influences. This can be for a number of reasons. Firstly, many of these traits are under the control of a large number of genes, some of which might be sensitive to different environmental factors such as day length, rainfall, soil *p*H and so on. Therefore a particular combination of these factors might affect markedly the expression of a trait. Secondly, as a large number of accessions are, in reality, populations (i.e. their genetic constitutions are not uniform and are likely to show variation in preliminary evaluation trials which are carried out under different conditions) there will be a shift in the mean values for particular traits. Sometimes this difference in performance is quite striking. Therefore the yield of a group of accessions might vary from year to year (depending on the prevailing conditions). So if a scale were constructed based on one year's results, it might be inappropriate in following years where the mean yield might be markedly higher or lower. Also, if a new group of accessions were received, their variability with regard to yield might be unexpectedly high, so much so in fact that the scale might be unusable. This is particularly true for accessions which have been collected from different habitats.

If the distribution is found to be multi-modal and data are found to be distributed throughout the entire range, a scale is constructed which covers the range perhaps using regular intervals. On the other hand, if the data are not distributed throughout the entire range and show quite distinct modalities, the descriptor might more appropriately be scored using a nominal scale.

A good example for this last case is flower colour: it could conceivably be scored on a continuous scale as the wavelength of light in nanometres. However, in practice it is recorded on a nominal scale (e.g. cream, yellow, pink, red). This is because it shows quite distinct modalities; these modalities are used to define the descriptor states.



4 Dealing with heterogeneous data

We saw earlier on that genebank accessions are commonly heterogeneous, that is, each accession is not genetically uniform and contains a certain amount of variation (see fig. 3). It is a tricky problem for the documentation specialist: how do you reflect this variation adequately in the documentation system? How do you score these descriptors?

First of all, do you need to document this heterogeneity? If there's no current or future need, it's a low priority so why bother? What sort of detail do you need the heterogeneity documented -very detailed, or just a note that the data are "variable"? What will you be doing with the data once documented – analyses, reports, nothing in particular?

If you anticipate doing a lot of analysis in the future, there's only one thing you can do in practice: RECORD THE RAW EXPERIMENTAL DATA (original observations).

If this is not possible for practical reasons, there are a number of other approaches you can take.

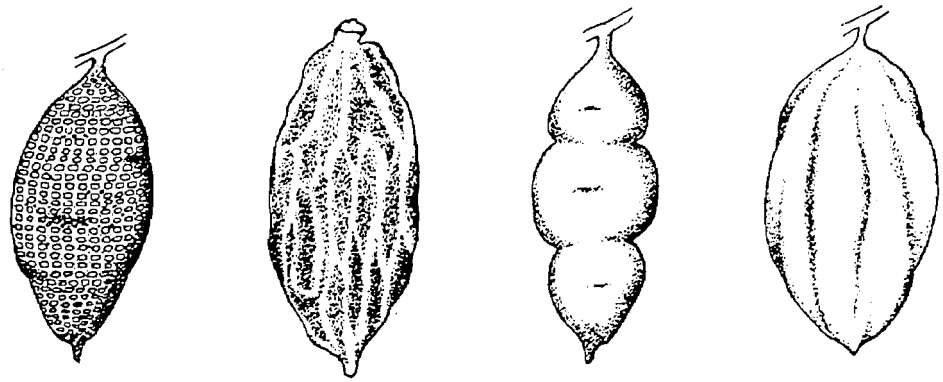


Fig. 3. Heterogeneity in storage root surfaces in sweet potato

4.1 Record the mean and standard deviation

This is probably the best approach to use for continuous scale data. It gives an idea of the mid-point of the range of values for a particular descriptor and the extent of variation shown in the sample. It is especially useful for normally distributed data as useful predictions can be made from these figures. However, if data show a multi-modal distribution these figures can be misleading.

4.2 Record the mean or most frequently occurring state

This is an easy way out because it ignores the problem entirely: all information on the extent of variation is lost!

4.3 Record the frequency of each descriptor state

If you were looking at flower colour of plants in an accession, with this approach you'd record that, say, 25% were yellow, 10% cream, 50% orange, 15% red and 0% purple. Genetic data are often of this form: you can deal with them in a similar way. For example, if you had carried out an isozyme study, you could record the percentage of plants in an accession showing each of the alleles at a particular locus.

Another method would be to list up to three codes in order of frequency but not quantify them further. In both cases, however, you might have problems in retrieving information from the system.

4.4 Record the range of variation

You could store plant height for a particular accession as "0.75-1.2m"; this would tell you something about the total extent of variation but nothing about where most values lie.

4.5 Score using a binary scale

This will indicate if an accession is heterogeneous but it will not be able to say how heterogeneous the sample is. You would then have to deal with the drawbacks of handling binary data: problems of data redundancy, of data recording and of information retrieval.

4.6 Record as "variable"

This is probably the easiest way out because it ignores the problem entirely: all information on the extent of variation is lost and there is no indication of where most values lie.

5 Where to go from here

From your analysis of genebank procedures in Chapter 4, you identified a number of descriptors which form the basis of the documentation system. You should now choose the appropriate scale for each of these descriptors, carefully considering the following points which have been raised in this chapter:

- The ease of data of recording
- The ease of data updating/modification
- The ease of flexible information retrieval
- The potential for data analysis/transformation

In the next chapter we'll be looking more closely at the design of manual forms for recording raw data and/or scoring descriptors and the use of the these forms to assist in genebank management.

6 Exercises

1. Which of the following descriptors can be scored using a continuous scale?

   a. Soil *pH*
   b. Collection source
   c. Plant height
   d. Monthly rainfall
   e. Harvest date
   f. 1000 seed weight
   g. Kernel type
   h. Accession number
   i. Foliage diameter
   j. Soil texture
   k. Topography

2. State whether the following descriptors can be scored using an ordinal scale or a nominal scale:

   a. Soil type
   b. Soil depth
   c. Soil drainage
   d. Flower colour
   e. Fruit shape
   f. Eating quality
   g. Disease susceptibility
   h. Leaf hairiness
   i. Evaluation environment
   j. Collection source

EXERCISES

3. For each of the following examples (a-t):

(i) Which descriptors can be scored on a continuous scale *and* an ordinal scale?

(ii) Which method would you choose and why?

a. 1000 seed weight
b. Accession number
c. Altitude of collecting site
d. Collection source
e. Foliage diameter
f. Harvest index
g. Harvest date
h. Kernel type
i. Leaf width
j. Moisture content (%)
k. Monthly rainfall
l. Number of flowers per plant
m. Plant height
n. Protein content (%)
o. Seed weight
p. Seed moisture content (%)
q. Seed viability (%)
r. Soil *p*H
s. Soil texture
t. Topography

4. The following nominal scale lists the possible sources for collecting maize. Convert them into a list of separate binary (presence/absence) observations. How are the data best expressed – as a nominal scale or as separate binary observations? Give reasons for your choice.

l=wild habitat
2=farm land
3=farm store
4=backyard
5=village market
6=commercial market
7=institute
8=other

5. What is wrong with the following ordinal scales? Correct them where appropriate.

a. Scale for "Plant height":
1=intermediate to tall (1.5-1.75m)
2=intermediate (1.25-1.5m)
3=short (0.75-1.0m)

ANNEXES

4=short to intermediate (1.0-1.25m)
5=tall(1.75-2.0m)
6=tall to very tall (2.0-2.25m)
7=very tall (>2.25m)
8=very short (<0.5m)
9=very short-short (0.5-0.75m)

b. Scale for "Presence of seed":
1=none
3=few
5=medium
7=abundant

c. Scale for "Density of flesh of ripe fruit":
1=very dense (very fleshy)
3=dense (fleshy)
5=intermediate
7=loose (crumbly)
9=very loose (spongy, large air cavities)

d. Scale for flower colour:
0=yellow
1=orange
2=red
3=dark purple
4=purple

e. Binary scale for "Type of fruit" in pineapple:
0=grouped fruitlets
1=isolated fruitlets

Chapter 6

Organisation of different types of data

Chapter 6 deals with the use of manual forms in recording and organising data and looks at the organisation and operation of a manual documentation system. When you have finished this chapter, you will be able to:

- Explain why it is not always possible to record data directly into a documentation system
- Discuss design considerations for manual forms
- Describe the composition and operation of a management file in manual documentation systems
- Describe ways of coping with feedback information and raw data

1 The use of manual forms

Many of the documentation procedures that you develop will rely on the use of manual forms for recording raw data or scoring descriptors even if you are using, or are planning to use, a computerized documentation system.

We saw earlier how helpful it is if the documentation is an integral part of the genebank procedure. Such integration is necessary to avoid the duplication of effort. You should therefore look at the genebank procedures to see how practical it is to record data *directly* into the documentation system. If direct data entry is *not* possible, it's usually for the reasons set out below.

1.1 Direct data entry is impractical

If you have a computerized documentation system and you are doing field evaluations or working in the seed store at -20°C, you can't really carry your computer around with you (and you probably wouldn't want to anyway!). In such cases you should take a notebook or a data logger with you which you can use to record data. You would later transfer these data to the documentation system.

1.2 Data analysis required before documentation

In some cases it's not possible to score the descriptors directly as you need to do some collation and analysis of the data beforehand.

1.3 Operational difficulties

If you are recording data from a chain of procedures it might not be possible to record the data directly for operational reasons.

For instance, if you are operating a computerized documentation system with one computer, you can imagine the situation where several people will be waiting to use the computer to record the data before continuing with the rest of the procedure. This would actually *slow down* the operation of the genebank procedure – clearly a situation to be avoided! In this particular case you would probably record the data on manual forms and enter the data in the system at a later stage, or – better still – have one person dedicated to data entry.

In a manual system, a form could be developed for a chain of procedures which can be used by different people. Problems would arise if people needed to record data on the form at the same time. This would slow things down. In this case you might have to divide the descriptor set into smaller but meaningful descriptor sets and develop separate forms for each of the smaller sets: the forms would then be used by one person at a time.

Table 1 will help you clarify which descriptor sets can be scored directly and which cannot.

Table 1. Possibilities for direct scoring of descriptors in a documentation system and the requirements for analysis

	DIRECT SCORING	ANALYSIS REQUIRED °
Registration	✔	
Passport	✔	
Characterisation		✔
Evaluation		✔
Seed handling		✔
Inventory		✔

2 Design of manual forms for computerized and manual documentation systems

It's important that any forms used, whether for a computerized or manual documentation system, are user-friendly to ensure data integrity. Typically, you can use two sorts of design of layouts: column layouts (for several accessions) and a page layout (for a single accession). Let's look at these in more detail.

2.1 Column layouts (several accessions)

Columns are very practical as they are easy to use. Each form should be:

- Easy to fill in
- Easy to update (where appropriate)
- Easy to read

The following points should give you ideas on how to design a page layout with columns.

2.1.1 Use pre-printed forms or ruled paper

Try wherever possible to use pre-printed forms when recording data – it will make the task so much easier. You can design these on a word processor if you have one.

If you cannot use pre-printed forms, use ruled paper. Don't try to write on paper with narrow ruled lines otherwise you will run into problems of space especially if you have large handwriting. As a result the form will be untidy and difficult to read. What is the best way to get round this problem? Don't use the narrow ruled paper in the first place – use paper with wider ruled lines instead. Also, draw the columns clearly with a fine point pen or sharp pencil and a ruler – don't use anything that comes to hand like a cigarette packet or the edge of a comb(!). Keep the columns parallel as you don't want them to get narrow at the bottom of the page.

2.1.2 Make each column wide enough

It's a great temptation to try to put as much data as possible on one page by having numerous, narrow columns on the page. Don't! It will make data recording and retrieval much more difficult: you will have difficulty writing in the narrow columns and difficulty reading your writing afterwards. There will also be the temptation to invent short-hand

codes at the time of recording so that you can comfortably fit the data into the column. This will be disastrous later on when you have difficulty in remembering what all the codes mean.

If you are having trouble with narrow columns you can:

1. Turn the paper around so that the top of the page becomes the side of the page – so-called landscape orientation (see fig. 1). This will give you more room.

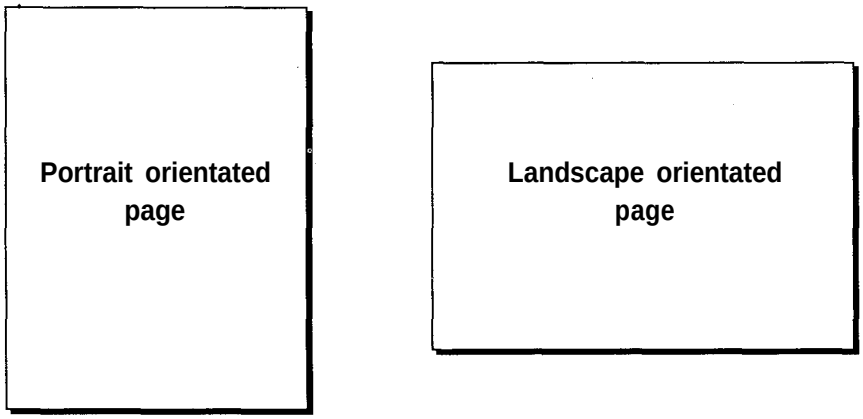


Fig. 1. Examples of portrait orientation and landscape orientation

2. Use wider paper, but not so wide that it becomes cumbersome.
3. Consider using numeric codes as described in Chapter 5 – if you're not using them already.
4. Ask yourself whether you are trying to record too much data on the form and if so, whether the data could be divided into smaller, meaningful sets that can comfortably fit on separate pages or forms.

2.1.3 Arrange the columns to help data recording and retrieval

You should arrange the columns in an order which helps you record and modify the data, particularly if you are transferring the data from lab books or other forms where the data have already been recorded in a regular order. Therefore use the same order as is in the lab books and on the documentation form. If you don't do this you will waste time reorganising the data each time you transfer them and you might make mistakes:

For manual systems: If you know that you will be modifying data at a later date (e.g. amount of seed in store), it's helpful if you adopt the convention of putting these data in columns near the right hand side.

You should also arrange the columns to help you retrieve the data. Therefore for all accession-specific data you should use the first column of all forms for the accession number. Where appropriate you could use the next three columns for the scientific name, batch reference and date of test (see fig. 2).

Accession number	Scientific name	Batch reference	Date of test	Test 1	Test 2	Test 3

Fig. 2. Example of columns arranged to help data recording and retrieval

If all the forms (where appropriate) are structured in this way, it will help you retrieve information from many different forms as you will know where to look.

2.1.4 Be consistent in the arrangement of columns

You should stick to the order of columns that you have chosen. You should not change the order unless there are problems with data recording-and retrieval. Alter the column width if it will help data recording, but if you change the order of columns it will cause confusion when you are browsing through several pages.

2.1.5 Include a “comments” column

If you have room, it's always useful to include a column where comments or remarks can be written. Put the comments column on the far right hand side of the page and make it fairly wide.

2.1.6 Give your form a title!

If you are designing pre-printed forms, don't forget to give them titles and a short explanation if necessary – you don't want people to use the wrong form. If you are not using the reverse side of the sheet, consider using the space for giving an explanation of the form and any coding systems used.

2.1.7 Try out several designs; determine which is most user-friendly

This is very important. Mistakes can easily creep into a documentation system at data entry so make sure that the form is as easy to use as possible. Try out a few designs and find out which design the users prefer. Dont choose a design which is awkward to use! The following examples of a bad page layout (see fig. 3) and a more usable page layout (see fig. 4) illustrate some points for you to bear in mind when designing your forms.

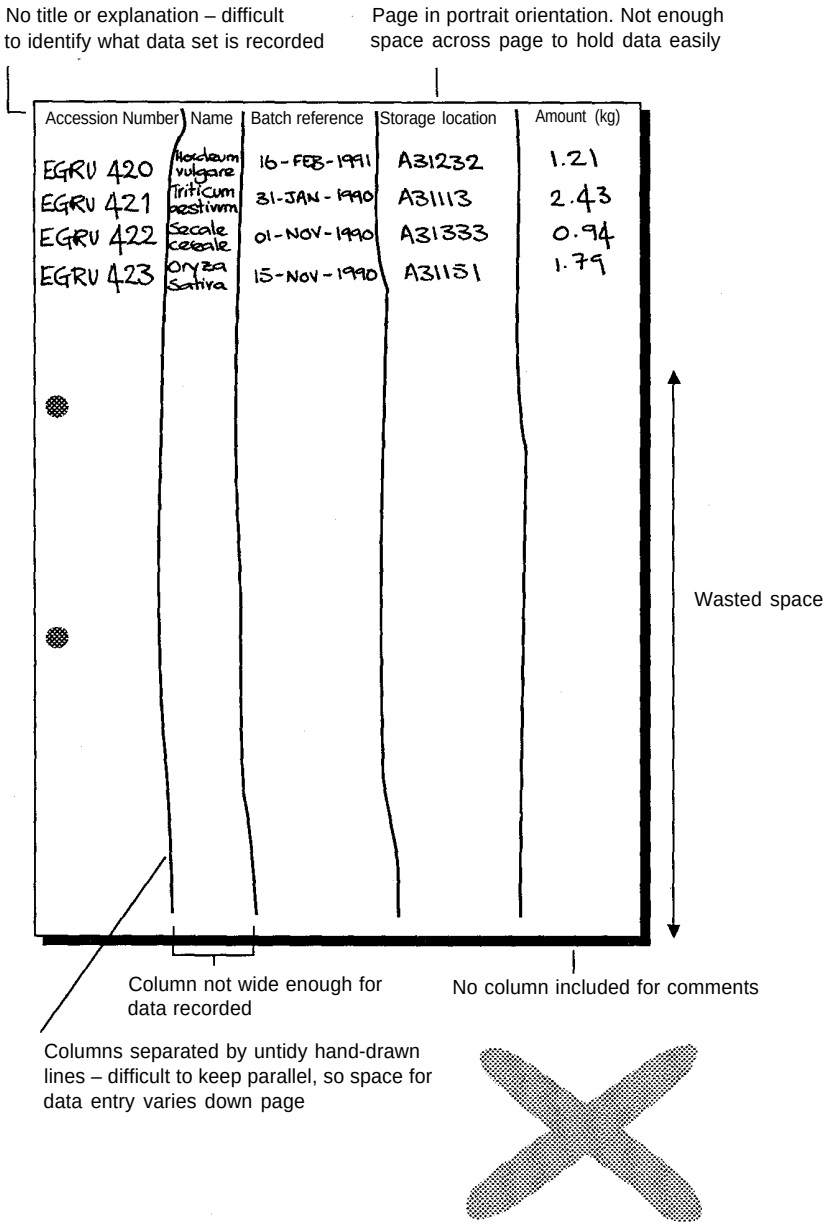


Fig. 3. Example of a bad page layout

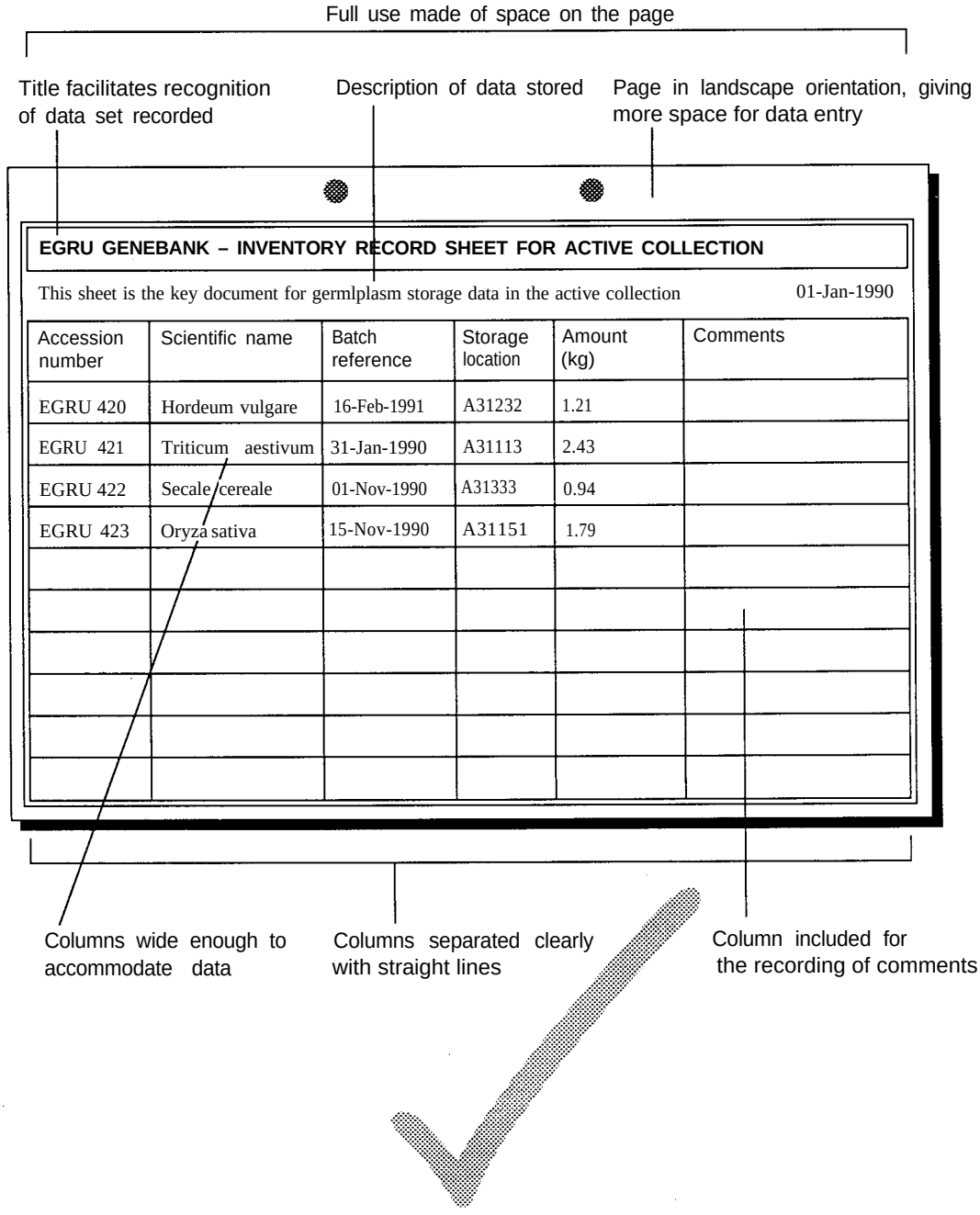


Fig. 4. Example of a good page layout that facilitates data recording and retrieval

2.2 Page layout (single accession)

You should use this sort of layout where you are recording a lot of data for a particular accession. A good example is a form used by plant collectors where many data and miscellaneous observations need to be recorded.

With a page layout you have considerably more freedom in where you can place descriptors on the page. With page layouts you can use:

- A mixture of columns or text boxes
- Multiple choice questions to assist the user
- Comments to assist the user

Don't get carried away with this extra freedom and end up with a form which is cluttered, confusing and not so easy to use. Be careful when designing the layout on a word-processor – it's sometimes difficult to know how practical certain designs are. Try out a few different layouts to find out which layout is easiest to use.

Many of the design considerations we discussed with column layouts also apply to page layouts, namely:

- Try to use pre-printed forms, otherwise use ruled paper
- Draw the columns or boxes neatly
- Make each column or box large enough to accommodate data
- Aim for consistency in different layouts
- Include space for "comments"
- Give your form a title and give indications for use
- Try out a few designs; choose the most user-friendly design

There are a few other points that should be considered for page layouts.

2.2.1 Make full use of the page

Make full use of the page – don't cram descriptors into a small area.

Decide how much room is needed for each item. If space is not limited, be generous – give each item plenty of space to allow easy recording. Where the data is recorded next to the title of the field, leave enough vertical space (as well as horizontal space) to record the data.

2.2.2 Use text boxes for emphasis only

Use text boxes for emphasis. Use them for emphasising descriptors that you will want to locate quickly such as the accession number. Dont use them routinely for all descriptors as it will make the page difficult to read. A few well placed text boxes are very useful in locating data at a later stage.

2.2.3 Use different styles of text sparingly

Some layouts look aggressive with large, bold, titles and excessive explanatory notes. Sometimes these forms have little room for data recording! If you are designing the form using a word processor, remember that you have different styles of text (e.g. **bold,** *italic*, underline). Use them sparingly and for emphasising important parts of the text.

As a general rule, do not mix the different styles together, use one style at a time.

3 Organisation of a manual documentation system

This section deals with the organisation and operation of a manual documentation system. Even if you are operating or plan to operate a computerized system, the section is relevant as you are bound use some manual forms in your documentation procedures which can later be used as a back-up for the main system.

The secret to designing a good manual documentation system is to design one which is:

- Simple
- Easy to understand – coupled with well defined documentation procedures
- Easily accessible – for the people who will use it

The stages that you should follow in building a manual documentation system are illustrated in fig. 5.

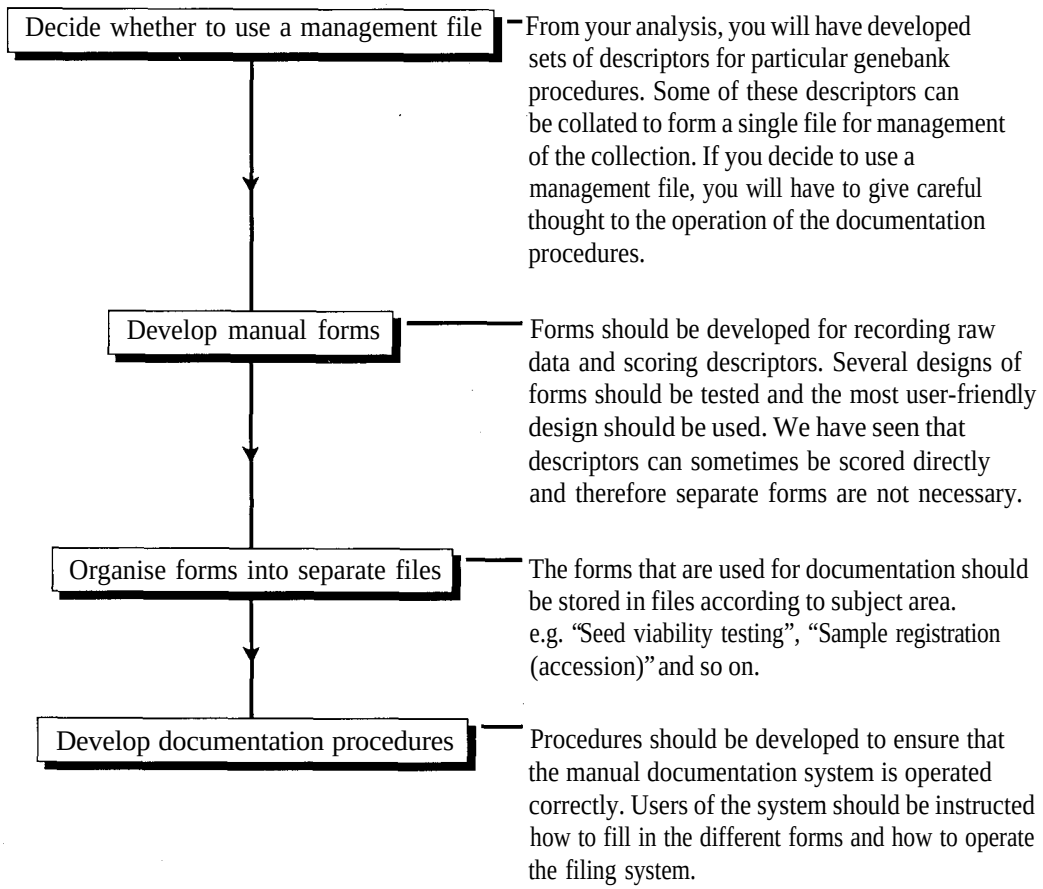


Fig. 5. Stages of building a manual documentation system;

3.1 Use of a management file

Genebanks which run only manual documentation systems commonly operate a management file which contains the essential management descriptors collated from the different genebank procedures. Such a management file might contain some or all of the descriptors listed in table 2. Further descriptors, including descriptors for field and *in vitro* collections can be found in Chapter 4.

Table 2. **List of possible descriptors for a management file (seed collections)**

DESCRIPTOR
Accession number
Scientific name
Batch reference
Collection type
Location in store
Total weight of seed
1000 seed weight
Minimum amount of seed allowed
Moisture content
Viability (%)
Date of viability test
Date of next viability test
Date of supply (distribution)
Amount of seed sent (distribution)
Recipient code (distribution)
(Other descriptors)

Forms can be designed which contain these and other descriptors. Since there are a comparatively large number of descriptors and potentially more than one batch reference for each accession, genebanks commonly use one form per accession (see fig. 6).

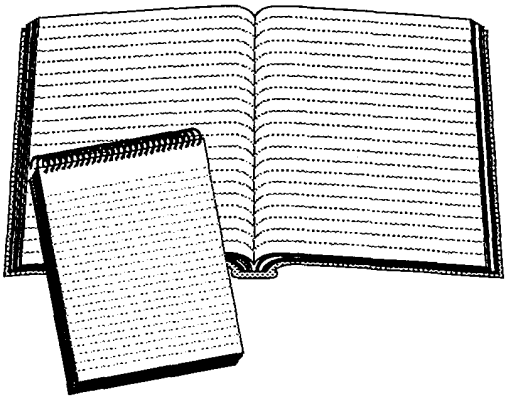
BASE COLLECTION			
Accession number		Origin	
Scientific name		1000 seed weight	
Acquisition date			
Donor ID number		Minimum amount of seed allowed	
	First batch	Second batch	Third batch
Batch date			
Amount of seed			
Moisture content			
Location			
VIABILITY TESTS			
Batch	Date of test	% viability	Date of next test

Fig. 6. Example layout of a management file (single accession)

It's very important that the management file is updated as a routine part of all genebank procedures which produce management data. In practice, the files concerned with seed handling will commonly contain the raw data, the descriptors being scored in the management file. This avoids any discrepancies in different parts of the documentation system as the descriptors are scored in one place only.

3.2 Materials

In a manual documentation system you will probably use some of the following materials, either writing on them by hand or sometimes using a typewriter:

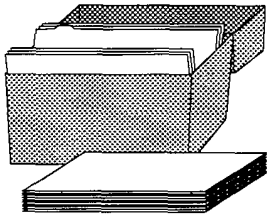


- Notebooks (hard back books, exercise books)
- File cards
- Loose leaf sheets

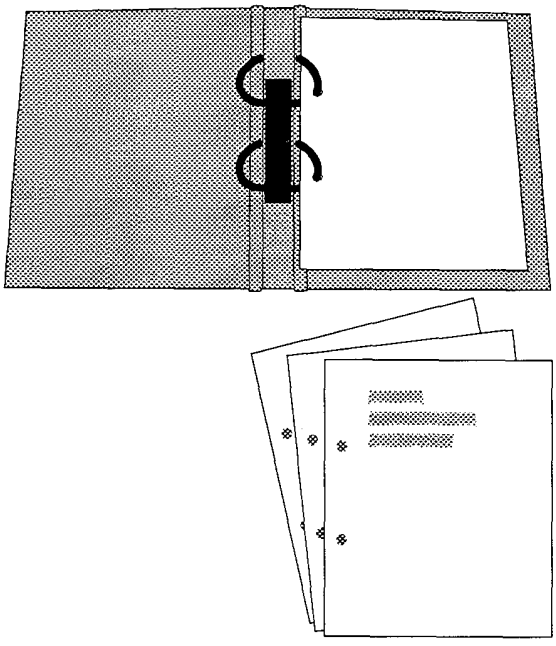
These different materials are sometimes called *storage media.*

You should not record the data in an arbitrary or random fashion on these storage media, you should design a page layout which will help you record and retrieve data for each descriptor set you have identified.

Notebooks can be used if the descriptors are the same for all accessions and the pages do not need sorting later on. If you organise data into columns in the book, you will be able to record (and retrieve) data in a straightforward way. You might consider using notebooks for registration or germplasm distribution. You probably would not use them for address lists unless you had divided the book into separate sections.



File cards are very useful for recording small amounts of data about a single item such as a genebank accession or an address. Use file cards when you want to sort the items later on and when you don't need to make quick searches for the same data over many germplasm samples. You can use them for address lists as these are always being updated but you probably would not use them for passport data (too much data) or characterisation data (need to make comparisons of many samples).



Loose leaf sheets are the most flexible but this depends on how they are being used. If you are storing data on one accession per sheet, they are like file cards in that you can sort them easily afterwards but with the added advantage that you can store more data. (Unfortunately, like file cards, they can be easy to mislay unless you are careful!). If you are storing data on more than one accession per sheet they are like notebooks (for ease of recording/retrieving data and direct comparisons between accessions) but with the added advantage that you can sort the sheets afterwards. For these reasons, loose leaf sheets are very widely used.

These features are summarised in table 3.

Table 3. Comparison of different storage media

	EASE OF SORTING	CAPACITY FOR DATA STORAGE	EASE OF COMPARISONS
Books: 1 accession per page	Difficult or impossible	Large	Poor
Books: > 1 accession per page	Difficult or impossible	Medium	Moderate to good
File cards	Good	Small	Moderate
Loose leaf sheets: 1 accession per page	Good	Large	Poor
Loose leaf sheets: > 1 accession per page	Difficult or impossible	Medium	Moderate to good

You can store the loose leaf sheets in the following:

- Files in a filing cabinet
- Ring binders
- Box file
- File wallets

You can store the file cards in a box or sliding trays. To simplify the discussion, we will call the different storage units *files*. Files should be clearly labelled so that their contents are obvious. If they are part of a series, this should be indicated. It's a good idea to include at the front of the file, instructions on how to use it and descriptions of any coding systems used. Also, all sheets stored in the same file should have the same format.

If possible, all files should be stored in the same place – you don't want to go from room to room to find different files. They should be easily accessible – you should not have to stand on a chair to get the files down. If space is limited, lab books and notebooks which are only referred to occasionally can be stored in less accessible places.

3.3 Making data easy to read

How can you arrange data on paper so that they are easy to read? It's simple – you need to develop a clear layout, i.e. how the data will appear on the page. Let's look at an example. Consider this address list of five International Agricultural Research Centres:

```
CIAT Centro International de Agricultura Tropical Apartado
Aereo 6713 Cali Colombia Phone: (57-23) 675050 Fax:
(57-23) 647243 CIMMYT Centro Internacional de Mejoramiento
de Maiz y Trigo PO Box 6-641 Mexico 06600 D.F. Mexico
Phone: (52-5) 954 2100 Fax:(52-5) 954 1069 CIP Centro
Internacional de la Papa Apartado 5969 Lima Peru Phone:
(51-14) 366920 Fax: (51-14) 351570 ICARDA International
Center for Agricultural Research in the Dry Areas PO Box
5466 Aleppo Syria Phone: (963-21) 213433 ICRISAT
International Crops Research Institute for the Semi-Arid
Tropics Patancheru P.O. Andhra Pradesh 502 324 India
Phone: (91-842) 224016 Fax: (91-842) 241239
```

Do you find it easy to read? Can you easily find the telephone number for CIP? What is the telephone number for CIMMYT? What would happen if you arranged it differently.

This format is easier to read but could still be improved further.

```
CIAT Centro Internacional de Agricultura Tropical Apartado
Aereo 6713 Cali Colombia Phone: (57-23) 675050 Fax:
(57-23) 647243

CIMMYT Centro Internacional de Mejoramiento de Maiz y
Trigo PO Box 6-641 Mexico 06600 D.F. Mexico Phone: (52-5)
954 2100 Fax(52-5) 954 1069

CIP Centro Internacional de la Papa Apartado 5969 Lima
Peru Phone(51-14) 366920 Fax: (51-14) 351570

ICARDA International Center for Agricultural Research in
the Dry Areas PO Box 5466 Aleppo Syria Phone: (963-21)
213433

ICRISAT International Crops Research Institute for the
Semi-Arid Tropics Patancheru P.O. Andhra Pradesh 502 324
India Phone(91-842) 224016 Fax: (91-842) 241239
```

ACRONYM	ADDRESS	PHONE	FAX
CIAT	Centro Internacional de Agricultura Tropical Apartado Aereo 6713 Cali Colombia	(57-23) 675050	(57-23) 647243
CIMMYT	Centro Internacional de Mejoramiento de Maiz y Trigo PO Box 6-641 Mexico 06600 D.F. Mexico	(52-5) 954 2100	(52-5) 954 1069
CIP	Centro Internacional de la Papa Apartado 5969 Lima Peru	(51-14) 366920	(51-14) 351570
ICARDA	International Center for Agricultural Research in the Dry Areas PO Box 5466 Aleppo Syria	(963-21) 213433	
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics Patancheru P.O. Andhra Pradesh 502 324 India	(91-842) 224016	(91-842) 241239

By arranging data carefully, it can be much easier to use. It may take up more space but it will save you a lot of time in the long run.

3.4 Importance of arranging data in a useful order

We have seen that you should arrange the data into a useful order, an order which helps you retrieve data and information. This order will not always be accession number order. It could also be by crop name, species name, date of trial, date of distribution, etc. In theory you could

sort on the basis of any descriptor but it would have to be both a *useful* order and a *practical* order. Remember that an order which is useful for data retrieval may not always be practical for data recording and data modification: you sometimes have to make a compromise. If an inappropriate order were chosen, you would spend too much time:

- Recording data in the system
- Keeping the system up to date
- Retrieving specific data

In a manual system you'll have to decide which is the most useful order at the beginning as it's more work if you change your mind later on.

3.4.1 Sort data before recording them

When recording data about several accessions on one form, it will help considerably if they are sorted in a sensible order prior to recording e.g. in accession number order. This will help all future updating of data and information retrieval.

3.4.2 Anticipate multiple records

Often you can record one genebank accession per line. So, on a form with 20 separate lines, there might be 20 separate genebank accession numbers (sorted in order, of course). However, there will be occasions where you will need to make more than one entry for each accession. For instance in an inventory file might need to store more than one batch of seed: sometimes these additional entries will be added at a later date. If you haven't reserved space for these additional entries, where are you going to put them? Your documentation system might start getting untidy.

Let's illustrate this point by looking at part of an inventory file in fig. 7. Where would you put a new batch for accession 100? It's a problem!

Accession number	Scientific name	Batch reference	Storage location	Seed weight (kg)
100	*Zea* mays	28-Jun-1986	A10101	2.50
101	*Zea mays*	28-Jun-1986	A10102	3.21
102	*Zea mays*	28-Jun-1986	A10103	2.35
103	*Zea mays*	28-Jun-1986	A10104	1.70

Fig. 7. An inventory file form with no space allowed for new batches

If you had reserved a couple of lines for new batches, it would be a lot easier. See fig. 8.

Accession number	Scientific name	Batch reference	Storage location	Seed weight (kg)
100	*Zea mays*	28-Jun-1986	A10101	2.55
		18-Jul-1990	A10213	3.72
101	*Zea mays*	28-Jun-1986	A10102	4.94
		18-Jul-1990	A10214	5.20
102	*Zea mays*	28--Jun-1986	A10103	3.90

Fig. 8. An inventory file form allowing space for new batches

Another point to take note of is that since the accession number and scientific name are the same for each batch of an accession, you only need to record these data for the first batch reference on this form. This will make the form easier to read.

3.4.3 Anticipate future data updating

Sometimes you will have to change the data that you have recorded on a form as a result of certain genebank operations. So, for the inventory file, you need have to change the "seed weight" every time you remove seed from the store. If you had a narrow column on the form you would have to erase the existing seed weight (with a pencil eraser or correction fluid) and write in the new seed weight. If you had to do this many times, it might become difficult to read the recorded value. To avoid this problem; you could make the column wider and simply cross out the old value and write in the new value (fig. 9).

Accession number	Scientific name	Batch reference	Storage location	Seed weight (kg)
100	*Zea mays*	28-Jun-1986	A10101	~~2.50~~ ~~2.28~~ ~~1.96~~ 1.43

Fig. 9. Columns with enough space for data which need updating

3.4.4 Use "flags"

If you are transferring data from notebooks to the documentation system, it's a good idea to get into the habit of putting a small tick ✔ or other "flag" in the top right hand corner when all the-data on that page of the notebook have been transferred. This is particularly useful when you are transferring data to more than one place in the documentation system as you want to know whether the data have been transferred. A flag can similarly be placed in the documentation system when all possible data have been entered on a particular page. That way, you will know whether any data are missing.

3.5 Coping with feedback infomation

In Chapter 3, we saw that data on genebank accessions do not only come from genebank operations. Other people may be using the germplasm as well and the chances are that they will be generating their own data. If work is being performed on behalf of the genebank, such as an evaluation trial, supply manual forms or computer formats for recording these data.

Other feedback data will be group data or accession-specific data published in reports, theses, literature references, conference proceedings or from unpublished observations from field trials or verbal communications.

Remembering that you want good data in your documentation system, you need to decide:

- The need for documenting these data
- The priority for documentation
- The information requirements from documented data

Commonly you should keep a reference to the source of the information (i.e. the literature reference) rather than try to incorporate detailed data into your system.

If there is a limited amount of feedback information, the simplest approach would be to have an *information sheet* for each accession. Use this sheet for all literature references concerning the accession in question and any vague comments ("high yielding variety") or specific information ("percentage leaf mineral content"). You might also attach to the sheet any letters or communications concerning the accession and store these in a file in accession number order. Make an information sheet only for accessions which have information, you don't need to make blank **ones.**

3.6 Keeping track of raw data

There will be occasions when you will want to look at the raw data or to find out the precise conditions of a characterisation trial. In these cases, you'll make a reference to the source of the raw data in your documentation system. This means that you'll have to keep the raw data, not throw them away!

How can you keep track of the raw data?

3.6.1 Keep a diary

Diaries or wall-charts (see fig. 10) are useful for planning your work and making the best use of your time. They will not form part of the formal documentation system as such but it will be a useful tool for planning and a later source of information. Of course, if you work in a large genebank, it would be difficult to keep a diary of everything that goes on in the genebank but in smaller genebanks, the diary can be a useful source of information later on.

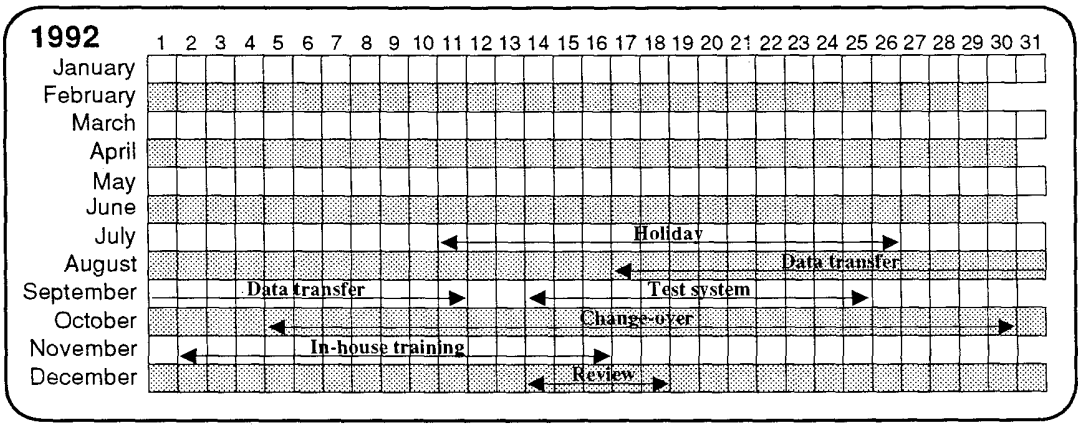


Fig. 10. Example of a wall-chart diary

3.6.2 Keep the original lab notebooks

Some people keep notebooks (or sheets) which they use to record the data on a day to day basis from their experiments, perhaps writing these data up more tidily later on (as reports) and/or formally recording them in the documentation system.

These notebooks should be kept, not thrown away. They should be filed according to subject and date so that the raw data can be located later on. The time period that each notebook or file represents should be clearly marked on the cover of the notebook/file (e.g. March 1990 – November 1991).

3.6.3 Make a reference to the source of the raw data in your system

For many genebank operations (regeneration, characterisations, evaluations) you can store the date of the tests in your documentation system so that you can locate the raw data at a later date. So, if an evaluation trial was started on 23-May-1990, the date of the trial is recorded in the documentation system and the original notebook is available for future reference.

4 Where to go from here

- Manual system only: decide whether to operate a management file. Determine how this file will be kept up to date in the different genebank procedures
- Draw up a list of forms which need to be developed
- For each form, experiment with a few designs to find the one which is most practical and user-friendly

In Chapter 10 we will return to ways of implementing the documentation system. Chapters 7, 8 and 9 deal specifically with computerized documentation systems.

5 EXERCISES

Exercises

1. Indicate whether the following statements are true or false:
   a. Manual forms are used only for manual documentation systems
   b. It is preferable to use pre-printed forms
   c. It is a good idea to design forms which can accommodate as much data as possible
   d. A manual documentation system can be used as a back-up system for a computerized documentation system
   e. It is possible to design more than one management file
   f. Consistency is ensured when management descriptors are scored in more than one place
   g. All manual documentation files should be sorted into accession number order
   h. Raw data do not form part of a documentation system and should be discarded
   i. Lab notebooks should be filed according to their accession numbers
2. List the occasions when it is not possible to record data directly into a documentation system.
3. How would you design a data entry form to make it user-friendly?
4. List the stages of building a manual documentation system.
5. What sort of data should a management file contain? Explain how you would operate the management file to avoid discrepancies with other files.
6. Why should data be arranged in a useful order in a documentation system? What are the consequences of using an unsuitable order?
7. How can you deal with feedback information in a documentation system?
8. Why is it a good idea to keep raw data? Discuss ways that you might use to keep track of raw data.

Chapter 7

Computer basics

Chapter 7 will introduce the basic features of computers – what they consist of, how they are operated and their use as tools for carrying out a wide range of activities. When you have finished this chapter you will be able to:

- List common microcomputer uses
- Describe a typical microcomputer set-up
- Define some basic terms used regularly when discussing computers
- Explain in general terms what a microprocessor does
- Explain the uses of different computer disks
- Describe the function of an operating system and two different types of user interface
- Discuss how microcomputers differ in microprocessor design and the operating system they use
- Be aware of whether or not software will operate in a given hardware set-up

1

Introduction

If you have never or only rarely used computers, one of the biggest problems to overcome when learning about them is the highly technical language that is often used. Sometimes it can be quite intimidating. The acronyms, abbreviations, strange words (some of which are quite funny), and strange concepts can seem like a lot of nonsense. You might ask yourself "What do I need to know to use a computer?"

Think about when someone is learning to drive a car. They are taught the basics of how to operate the vehicle, not how the internal combustion engine works. Being unable to replace the engine doesn't make someone a bad driver. But knowing some basic information about how the car works would help if, say, the engine failed to start.

You can think in the same way about computers. You have to learn how to use them for specific applications and generally know how to "drive" them. But they can be valuable tools without you knowing all there is to know about them. The more that you use computers, the more knowledge you'll gain. But you shouldn't acquire this knowledge at the

expense of losing sight of why you acquired the computer in the first place – as a tool.

This chapter introduces some basic information you need to know to be able to use computers, focussing on the kind of computer most commonly used in the laboratory, office and home. These are called *personal computers* (usually abbreviated to *PCs)* or *microcomputers.* They are sometimes referred to as *desktop* computers as they can be used on a normal office desk.

The basic features of a microcomputer set-up are discussed along with how (and why) microcomputers differ from one another. As detailed operating instructions differ with the various types of microcomputers, these are not covered here.

2 Computer technology

Computers are being used increasingly in everyday situations. They are becoming an integral part of the way we live. One of the main reasons for this is the rapid technological advances in recent years that have enabled the electrical components used in the manufacture of computers to be mass-produced very cheaply.

2.1 Chips and microprocessors

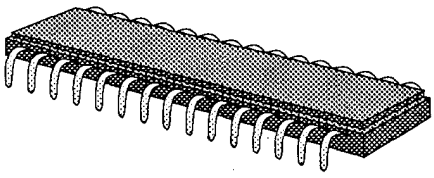


Fig. 1. A typical microchip

Silicon chips, microchips or just *chips* are thin slices of silicon (around 1mm thick) which contain several thousand electrical components, so small that you would need a microscope to see them. They are very powerful devices which have an enormous number of applications in the electronics industry.

A *microprocessor* is a very powerful chip which controls the other components of a given system. The components controlled by a microprocessor might be different chips or separate pieces of electrical equipment. For example, a microprocessor working with other components will perform the calculations on a hand held calculator, set the exposure

level on a camera and control the conditions in an electrophoresis experiment. There are literally thousands of situations where microprocessors are used. Does this mean there are thousands of different microprocessors, one for each situation? The answer is no.

A microprocessor has to be instructed what to do and how to operate. These instructions are contained on a separate chip. This means that the same microprocessor with different sets of instructions can be used for many different applications. Despite this, there are numerous designs of microprocessors, some more powerful than others, which are used for quite different purposes. For instance, a microprocessor used for aircraft navigation is likely to be more powerful than one used for alarm clocks.

2.2 Computers as tools

In the examples listed previously, the microprocessor is dedicated to one specific task. This is not true for the microprocessor in a microcomputer. It can be made to perform a variety of different tasks if given the relevant instructions. This means that the computer can be used for many different applications including: database management, word processing, accounts management, desktop publishing, graphic design and even for sending faxes. The computer is therefore a very flexible tool.

Since computers can be a valuable part of a genebank documentation system, we need to study them more closely,

3 Features of a typical microcomputer set-up

Definitions:

HARDWARE

The name given to all the physical items which comprise a computer.

PROGRAM

The name given to a series of logical instructions to the computer which perform a particular task.

SOFTWARE

The general name given to a program or group of programs which perform a particular task e.g. software for database management, software for word processing.

Note: The terms software and program are often used interchangeably.

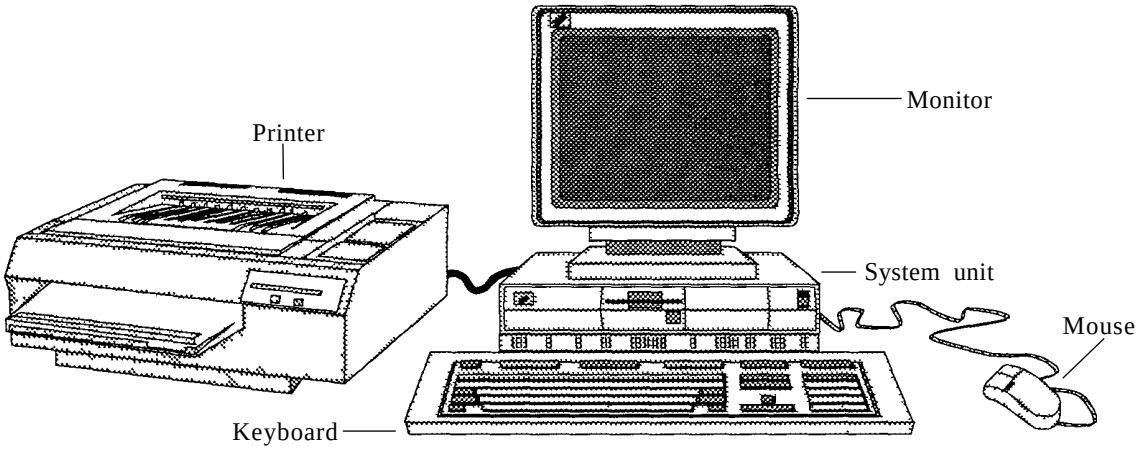


Fig. 2. A typical microcomputer set-up

A standard microcomputer hardware system is illustrated in fig 2. Typically it comprises a monitor, keyboard, a system unit containing the microprocessor, a printer and often a *mouse.*

3.1 Computer keyboard and monitor

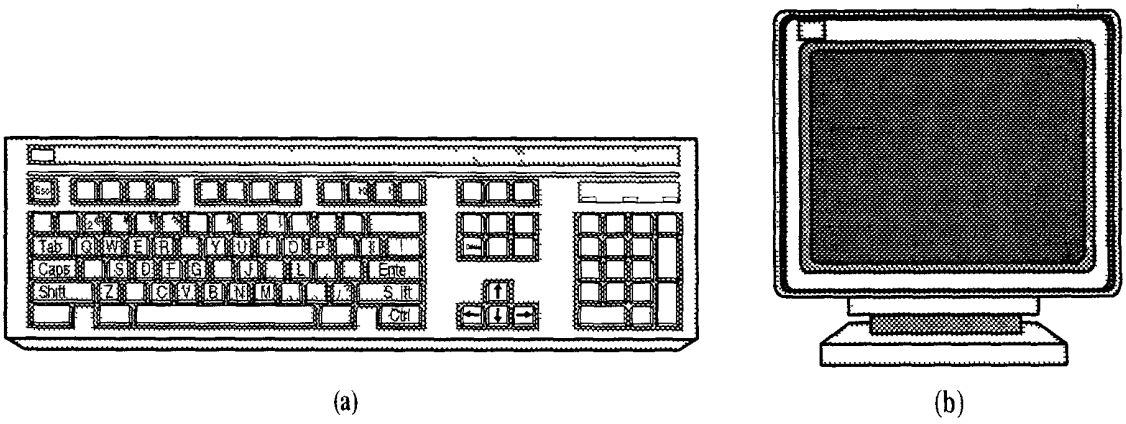


Fig. 3. (a) Typical computer keyboard. (b) Typical computer monitor

There are two ways of giving instructions to a microcomputer. The first is to type instructions on a typewriter-like ***keyboard*** (see fig. 3a). All microcomputers have a keyboard and a ***monitor*** (or visual display unit)

which displays what has been typed in (see fig. 3b). Instructions could be to start running some software. Or if software is already running and a document or file is being produced, the monitor displays what has been typed.

There are many different types of monitors available which vary in terms of the resolution capabilities, their hardware and software requirements and, of course, their price. From the lowest resolution to the highest resolution, these types of monitor include: Colour Graphics Adapter (CGA), Enhanced Graphics Adapter (EGA), Video Graphics Array (VGA) and Super VGA. When choosing an appropriate monitor there are **many** technical considerations to be addressed. For instance, if you were doing a lot of work with computer graphics, you would need a monitor with a high resolution such as Super VGA. Also, **some** specialised applications require the use of touch sensitive screens to enter information – the user simply touches an area of the screen. In all cases you should try out the monitor with the hardware/software combinations you will be using to check the quality of the display.

3.2 Computer mouse

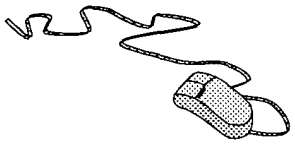


Fig. 4. Typical computer mouse

The second way of inputting instructions is by using a computer mouse. Many, but not all, microcomputers have a mouse attached. The mouse is a small, hand-held plastic box with one or more buttons **on** top, connected by a cable – or by radio signal in which case **no** cable is needed – to the system unit or sometimes to the keyboard (see fig. 4).

When the mouse is moved on a flat surface, usually an area next to the keyboard, there is a corresponding movement of a pointer on the screen. The pointer is often an arrow. Instructions are given by pointing at, clicking on and dragging objects displayed on the screen. Very little typing is involved. Graphic programs are usually difficult to operate without a mouse.

3.3 System unit

The brain of any computer is a microprocessor inside the system unit (see fig. 5.). This essential microprocessor is called a *CPU* or *central processing unit.* It controls the other components of the system, including other chips.

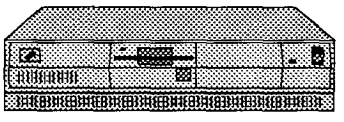


Fig. 5. A typical system unit

Special chips called *memory* chips are connected to the CPU. The computer memory is the computer's electronic working space. The size of working space affects the size of program a computer can handle and the speed a program can run.

A computer has different types of memory which are used for different purposes. *Random access memory* or *RAM* is the part of the memory usually used when working with programs or files. It is used for information which is referred to at random. Information contained in the RAM is lost when the computer is switched off.

The second type of computer memory is *read only memory* or *ROM.* As the name suggests, this type of memory contains information which can be read but not changed. The instructions a computer needs in order to start working when the power is switched on are contained in ROM. Information contained in ROM is not lost, even when the computer is switched off.

The system unit also has sockets to which you can attach the keyboard, monitor, printer and other equipment.

3.4 Computer disks

Computer disks are storage devices. You can store programs and your valuable data on them.

It's rather like using cassettes in a cassette/radio; you can record music on a cassette and play it back later on. If you decide you don't want the music any more you can erase it by recording new music on top of the old music. Similarly, the computer can be instructed (using the keyboard or mouse) to use (or play back) the programs and/or data stored on the disks. If you decide you don't want to store specific programs or data any more, you can instruct the computer to erase them.

There are two sorts of computer disks, hard disks and floppy disks. These are discussed in the following sections.

3.4.1 Hard disks

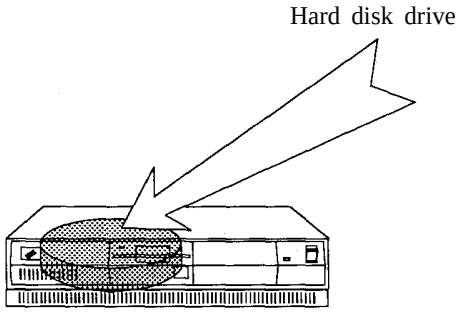


Fig. 6. Schematic representation of a computer hard disk

Most computers have what is called a *hard disk*. Most hard disks are stored permanently inside the system unit and are not removable. They normally have a large storage capacity and you'll commonly use them to store all your programs and data. Hard disks work faster than floppy disks when storing and retrieving data or programs.

3.4.2 Floppy disks

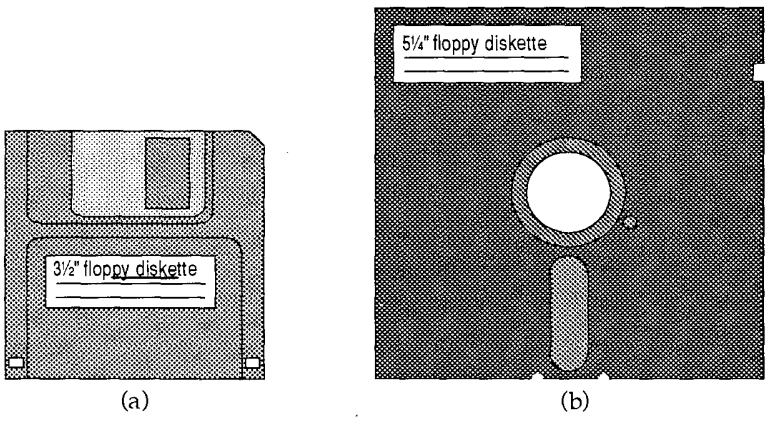


Fig. 7. Two common floppy disk sizes: (a) 3½ inch (b) 5¼ inch

The second sort of disk is called a *floppy disk.* There are two sizes commonly used, 3½" and 5¼". Different types of floppy disks can store varying amounts of data, but they all have a much smaller storage capacity, often *one hundred times smaller,* than hard disks. They do however have the advantage of being very portable so you can carry your work around easily. Floppy disks are commonly used for putting new programs or data onto the hard disk or transferring programs/data between different computers. You'll find them very useful when you exchange data with other genebanks.

Floppy disks store data through the action of a magnetic head writing on to the disk surface. You must take care to not expose any disk to a magnetic field as the data stored can become corrupted or erased. Magnetic fields can result from radiation emitted from behind computer monitors, magnets and other electrical appliances. Floppy disks are not damaged when exposed to x-rays.

Because floppy disks, especially 5¼ size, are made from a light and flexible plastic, they should be packed in a cardboard or other rigid box before transportation or posting to avoid damage.

3.4.3 Formatting disks

Any disk must be prepared before it can store information. This preparation process is called *formatting.* It organises the surface of the disk to allow information to be stored-and read. Formatting will erase any information which is already stored on a disk. You should consult your main computer manual for instructions on how to format disks.

3.4.4 Disk drives

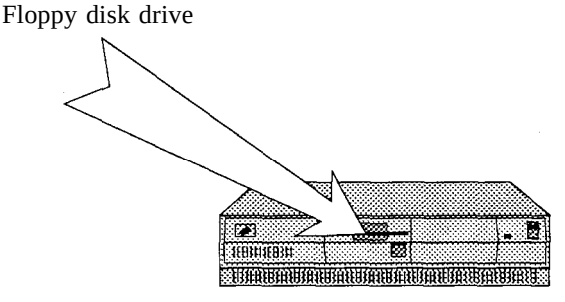


Fig. 8. Location of floppy disk drive on a typical system unit

The system unit contains spaces for computer disks. These spaces contain devices which copy information onto and retrieve information from the disks. These devices are known as *disk drives.* Floppy disk drives

are situated in visible slots on the system unit. However, the hard disk drive is concealed inside the system unit.

3.4.5 CD-ROMs

CD-ROM (compact disk read only memory) is an adaptation of audio technology which uses disks identical in appearance to audio *compact disks* or *CDs.* These disks are also referred to as *optical disks.* Using a CD-ROM is a compact way of storing large amounts of data. It allows the user to access large data files (often literature databases, images, illustrations and complete text works) which would take up too much space to keep permanently on a standard microcomputer. The actual CD is very portable, but in order to read the data, a different type of drive, a CD-ROM player (see fig. 9), must be attached to the system unit.

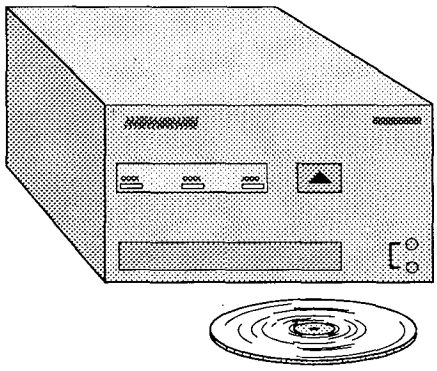


Fig. 9. A typical CD-ROM player and optical disk

Recent developments in CD-ROM technology have made it possible to both read data from and write data to optical disks. One example of this technology is a device known as a *WORM (write once, read many)* optical disk, which allows you to store data once, but read it as many times as you wish. This latest technology is not yet in widespread use. Until it becomes more common, most available CDs will allow you only to look at the stored data, rather than change existing data or add new data.

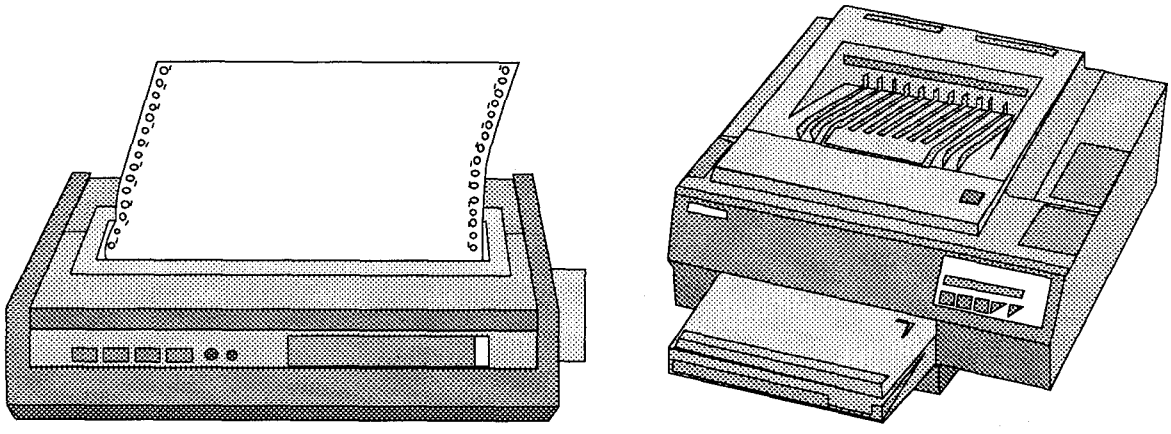
3.5 Computer printers

Frequently you will attach a printer to the computer for printing letters and other documents. There are many different types and models of printers available and the most common types are given in the following list.

1. *Daisy Wheel Printer:* This is a very basic type of printer, which generates characters in the same way as a typewriter – by striking an inked ribbon with a moulded character. This type of printer can be very noisy.
2. *Dot Matrix Printer:* This is another basic model of printer, which generates characters by electronically manipulating a matrix of tiny needles to form the required characters. The quality of characters generated by these printers is not high, but they are relatively inexpensive, fast and versatile. They can also be very noisy (see fig. 10a).
3. *Ink Jet Printer:* This is a "non-impact" type of printer. Electronic commands instruct ink drops to be squirted onto the paper to form the characters. The output quality is much better than with dot matrix printers and they are much quieter, but also are more expensive.
4. *Laser printer:* These printers use laser technology to produce high quality images for both text and graphics. Most offer a wide selection of fonts. They are more expensive than many other types of printers (see fig. 10b).

Dot matrix and daisy wheel printers are the most basic types of printer, but will be quite adequate for printing simple reports. However, if you have access to another type of printer such as an ink jet printer or a laser printer, you will find them more flexible for printing reports and letters and they have the potential to look more professional.

Different printers can use different languages to describe the way images or text appear on a page. These are called *Page Description Languages.* Some printers are able to print pages in different languages. PostScript™ is an example of a commonly used page description language.



(a) (b)

Fig. 10. Typical computer printers: (a) Dot-matrix (b) Laser

3.6 Other equipment

Other equipment can be attached to the computer which can enhance its operation and usefulness. Some of these are discussed in the following sections. Remember that computer technology is constantly evolving so try to keep yourself informed with the exciting new developments by reading computer magazines and journals.

3.6.1 Uninterruptable power supplies (UPS)

Power cuts are a source of worry for people working with computers because when they happen, work can be lost, files become damaged and data corrupted. You can protect your system from power cuts by using an *uninterruptable power supply* (UPS). A UPS acts as an interface between your computer and the main electricity source; when the UPS detects a drop in power or a loss of power altogether, it immediately steps in and supplies power from its own batteries. It cannot supply power indefinitely but will give you enough time to save your work and exit the application you are using. A UPS is therefore highly desirable if power cuts do occur from time to time at your place of work.

3.6.2 Voltage regulators

A voltage regulator is designed to protect your system against any fluctuations in the power supply such as a surge or drop in power. Such fluctuations can have disastrous consequences ranging from erratic behaviour of the equipment (leading to data loss) to widespread and irreparable damage to the system which might happen if the building were struck by lightning. A voltage regulator is therefore highly desirable if power fluctuations do occur-at your place of work. You-can often buy a UPS and voltage regulator combined in the same unit – check with your computer dealer for more information.

3.6.3 Modems

A modem (see fig. 11) converts signals generated by a computer into a form suitable for transmission over a telephone line. A modem at the receiving end converts the signals back for use by the receiving computer. A modem therefore allows communication between computers. Modems are commonly used to access remote computers which contain large databases (e.g. literature databases) or specialised software (e.g. for molecular biology). The remote computer might also have an electronic mailbox facility which allows you to send and receive electronic messages all over the world. Also, specialised modems (so called fax modems) allow you to send and receive faxes.

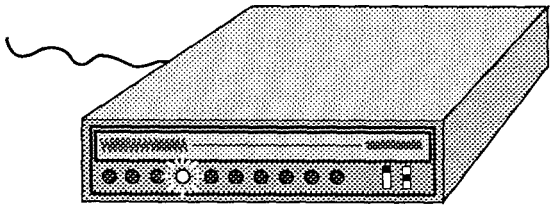
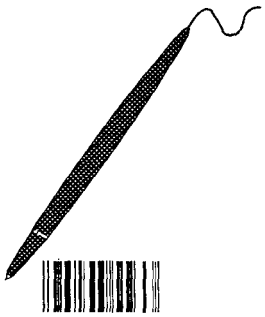


Fig. 11. A typical modem

3.6.4 Scanners

In the same way that a photocopier makes an image of a page and prints it, so a scanner can make an image of a page and store it as a computer file. Scanners are widely used in scientific work in the analysis of electrophoretic gels. In the office, scanners together with specialised software, are useful for reading text from a page and converting to computerized text thus avoiding tedious retyping.

3.6.5 Pen input devices



Pen input devices allow a direct but manual way of entering data or images into the computer. One common way is to use a *graphics tablet.* This is a flat plate on which the paper containing the image is placed; using a special pen, a series of positions on the paper can be marked and these are fed directly into the computer. This process is commonly called *digitisation.* This has great application when dealing with data from maps.

Another pen input device which has great application is the use of a light pen (or bar-code reader) to input data from bar-codes. As we saw in Chapter 5, bar-codes are being increasingly used for inventory management in genebanks and if the technology is available to you, consider using these to assist you in your work.

3.6.6 Additional storage devices

It's possible to buy additional disk drives for your computer, from conventional 3½" and 5¼" disk drives to removable cartridges which have a similar capacity to hard disks. You can also buy devices which use magnetic tape for backing up your entire hard disk – so-called *tape streamers* – ideal for archiving large data files and as a safeguard against potential loss.

4 How a microcomputer works

We've looked at the different features of a typical microcomputer and what the microcomputer can be used for. We've seen that the microcomputer consists of:

- A CPU which controls other components in the system (essential)
- A keyboard for entering instructions (essential)
- A mouse for entering instructions (optional)
- A monitor for displaying results (essential)
- One or more disks to store programs and data (essential)
- A printer for printing documents (highly desirable)
- Peripherals attached to the computer which can be operated by the computer (optional)

We've discussed how a microcomputer will do nothing unless it is instructed. So, how do you make a microcomputer work? How do you tell the microcomputer to operate as a word processor, a database manager and so on?

4.1 How a microcomputer starts up

The *boot sector* is the area of the disk which is referenced each time the computer starts up, for instructions on how to operate. When a computer is switched on, it first looks to the floppy disk drive to retrieve start up instructions. If the floppy disk drives are empty, the computer looks to the boot sector of the hard disk for its start up instructions.

4.2 Operating systems

When you first switch on a computer, it automatically 'runs' an important program called the *operating system.* This stays running the whole time the computer is switched on. The operating system controls the operation of programs and communication with the keyboard, monitor, mouse, disks, printers and any other electrical equipment attached to the computer. Without an operating system, you couldn't tell the computer to do anything. It is the interface between the user and the computer.

The operating system enables you to:

Operate programs (e.g. word processors, database managers etc.)

Organise your work (e.g. add, modify or erase any programs or data stored on the disk)

With a little more experience, you can also begin to use the operating system to customise the way that the computer operates and the way the information is displayed (e.g. the colour of the screen).

There are two main ways a user can interact with the operating system. This depends on the type of user interface.

1. *Command Line User Interface.* These systems require you to type instructions at the keyboard. To use these sorts of systems, you need to know some basic instructions. An occasional user can find this rather difficult as there are over 50 different instructions commonly used. An example of this type of operating system is MS-DOS® (see fig. 12).

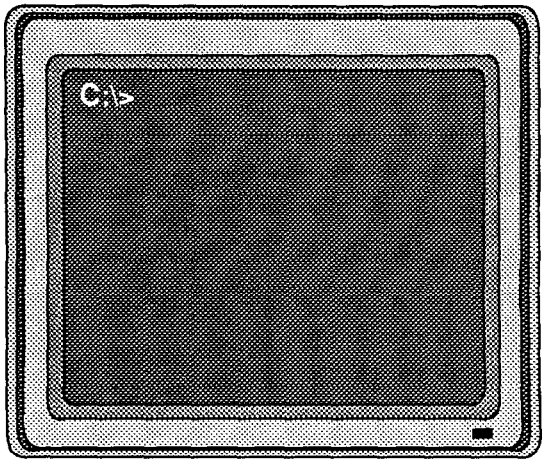


Fig. 12. Command line user interface displaying the screen prompt for the MS-DOS® operating system

2. *Graphic User Interface.* These systems accept instructions via menus and symbols selected from the computer screen (see fig. 13). The mouse is usually used for selection, but sometimes users find it faster to use the keyboard. Menus and symbols represent instructions to the computer, programs and data files that you have stored on your disks. Generally, graphics oriented operating systems are much easier for occasional users to operate. By pointing at and clicking on the symbols, you can explore the operating system and discover for yourself how to carry out tasks, with very little training. Examples of this type of operating system are Microsoft Windows™, and the Apple™ Macintosh™, operating system.

You need to know the basics about an operating system before you can start using a computer. But don't try to learn everything, it is very time-consuming and isn't necessary.

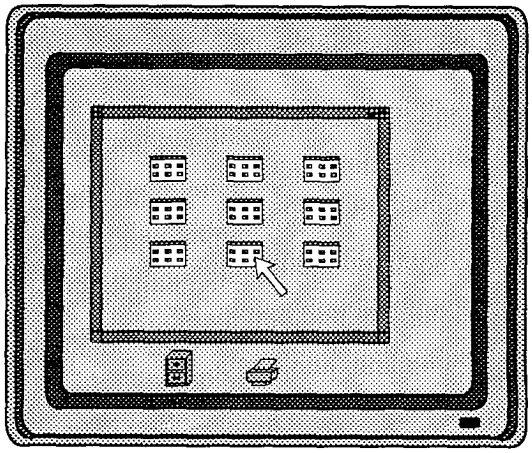
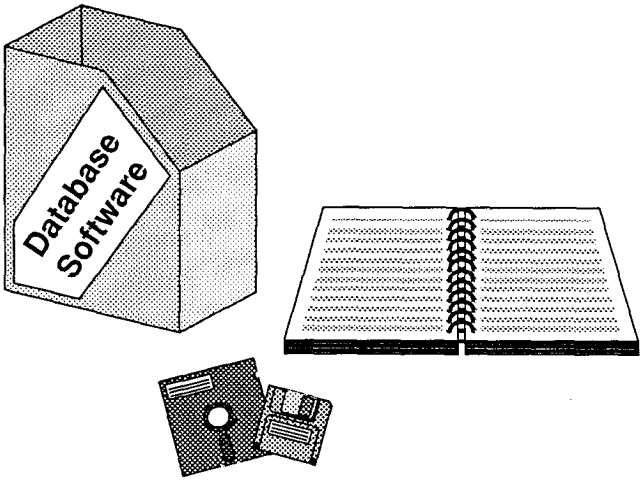


Fig. 13. Graphic user interface operating system

4.3 Software



How does the computer interpret your commands input via the mouse or keyboard into instructions to perform a specific task?

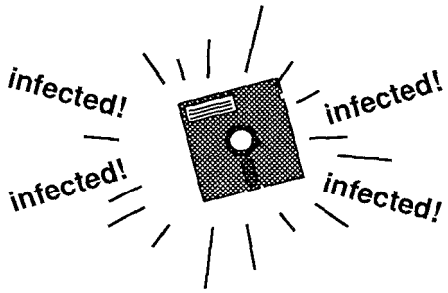
Software is a program or a set of programs which instruct the computer to perform specific tasks. It is the software which translates the commands you input into a language that the computer can understand.

Luckily, you are not expected to write all the programs yourself. There are many sophisticated general purpose programs which are widely available. For example, you can buy software for word processing, database management, accounting, project management, graphics design, desktop publishing and virus eradication and detection.

All commercial software that you buy will come with a detailed manual giving you instructions on operation and usually some tutorials to enable you to practice what you have learned.

Database management software can be a little more complicated to operate than other software. You often have to spend time learning the database language to make full use of the software.

5 Computer viruses



Computer viruses are undesirable programs which spread between computers, usually via floppy disks or over networks, and often cause damage to computer files. Like biological viruses, their effects can vary in degree. Sometimes they can be irritating but harmless. Sometimes they can have devastating effects, destroying your work and disrupting the operation of your computer.

Computer viruses can cause some or all of the following effects:

- Permanent damage to data files and program files
- Erasure of large areas of hard and floppy disks
- Slow operation of the computer
- Unexpected and erratic behaviour of the computer
- Changes in operation of peripherals (e.g. modems, printers)

5.1 Classification of computer viruses

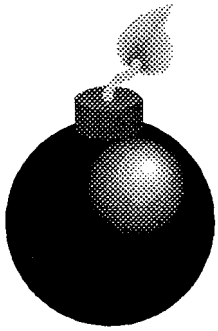
Computer viruses can be classified in a number of ways. The most useful way to classify them is according to which part of the system they attack. Two such classes can be identified: file viruses and boot sector viruses.

A *file virus* infects files, usually program files. When an infected program file is opened, the virus is activated and remains active while the program is used. Often it will spread to other program files. If the operating system files become infected, the virus will remain active the whole time the computer is running.

A *boot sector virus* infects the boot sector of hard or floppy disks. If this area becomes infected, the virus will be activated each time your computer starts-up. These viruses can spread very easily and very quickly as they are carried by the operating system of the computer, which is always running. Your computer can become infected simply by listing the files on an infected floppy disk.

Other types of viruses to be aware of are:

- *Worms:* These are viruses which infect computer networks
- *Bombs:* These viruses jump into action (or "detonate") on a particular date



- *Trojan horses:* These are not viruses in the strict sense. They are seemingly innocent programs (e.g. games programs) which cause severe damage to your computer disks when you try to operate them.

Viruses are a cause for concern and you'll want to be sure that none can get onto your system. There are a number of steps you can take to keep your system virus-free. These are covered in Chapter 10 on *Implementation and Maintenance of the System.*

6 Differences between microcomputers

To answer the general question: "Are all microcomputers the same?" we have to look a bit closer at two fundamental features of the microcomputer; the microprocessor and the operating system.

6.1 Microprocessor type

We saw earlier on in this chapter how at the heart of every microcomputer was a very special chip called a microprocessor which:

1. *Works with* other electrical components (including other chips)
2. To a large extent *controls* the other components and anything else attached to the computer. The design of this microprocessor is fundamental to the whole working of the computer.

There are different *families* of microprocessor, called families because they contain microprocessors based on the same original design. Every so often a new generation of microprocessors is developed with improved technical performance. These will still work with software written for older microprocessors. But some software written for the newer generation of microprocessor is written to take advantage of the new capabilities and therefore will *not* work with older microprocessors.

An example of a family of microprocessors is the Intel® family: 8088, 80286, 80386, 80486 (arranged in order of generation and increasing processing power). This family of microprocessors forms the basis for the popular IBM® PC series of microcomputer. Another example is the Motorola 680X0 series of microprocessors which are used in Apple™ Macintosh™ computers.

There are several families of microprocessor. Software written for one family of microprocessor will *not* work for another family of microprocessor. However, it's possible (in theory, if not in practice) to carry a program from one machine to another. In order to do this, it must be translated to work with the different microprocessor and operating system.

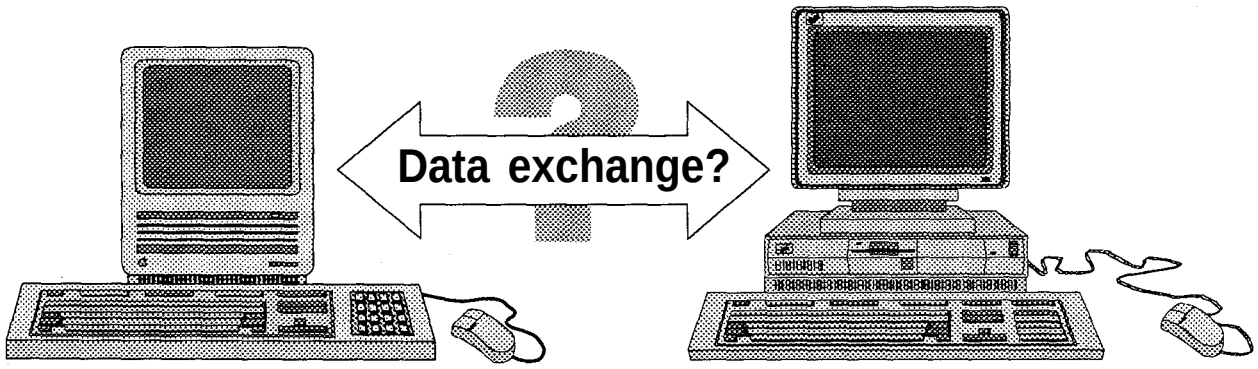
6.2 Operating system types

The operating system is designed to operate with a particular microprocessor or family of microprocessors.

We saw earlier that there were two types of operating system interface; the command line user interface and the graphic user interface. Sometimes you can choose to use either type of interface on your microcomputer but you can only use one at a time. Certain software will work with only one type of operating system. However, some software has been adapted to produce separate versions that will work with different operating systems.

Software manufacturers are always improving the operating system and from time to time release updated versions. Software written to operate with the older versions of the operating system will still work with the new version. However, some software written for the new version of the operating system is written to take advantage of the new capabilities and features and therefore will *not* work with the older operating systems.

Floppy disks are formatted for a particular operating system. This means that you can not always operate the same floppy disk in different microcomputers. It depends on which microprocessor/operating system combination the programs on the disk, or programs that produced the files on the disk, were designed to work with. Recent technological advances have made it possible for some microcomputers to use information recorded by a different microprocessor/operating system combination to their own. However, this is not yet commonplace.



6.3 The disk capacity

The *disk capacity* indicates the amount of data a disk can store. Disk capacity depends on both the storage properties of the magnetic surface of the disk and the way the disk is formatted, it is not directly related to the physical size. Most floppy disk drives on computers in current use can access data on floppy disks with different storage capacities. However, some older models of computer can sometimes only access disks with lower storage capacities. Consult your computer manual for more information on whether different capacities of disk can be used in your floppy disk drive(s). This will be of significance when you are exchanging data files between microcomputers, if drives on the different computers are not identical in this respect.

6.4 Other technical considerations

Some software may require, in addition to a specific microprocessor/operating system combination, additional power in the form of:

- A large memory
- A particular type of monitor
- A high storage capacity hard disk
- A computer mouse
- A particular type of printer
- Other equipment attached to the computer
- A math coprocessor – a microprocessor which works in conjunction with the CPU, which handles certain tasks (especially mathematical calculations) more efficiently

So, when you buy software look out for all these requirements. After all, you don't want to waste money on software that you can't use!

7 Safety precautions

There are a number of safety precautions you should take when operating your computer. Like other electrical equipment, computers are sensitive to the environment in which they are working and any adverse conditions can lead to eventual hardware failure which could cause loss or corruption of your valuable data. Many of these safety precautions are given in the handbook which is supplied with your computer but we can still list some general considerations.

7.1 The working area

- Keep your computer on a flat and sturdy work surface which has enough room to accommodate any peripheral equipment.
- Make sure there is enough space to allow adequate air circulation so that the equipment does not overheat.
- Make sure you have enough room to place any books or papers that you need to consult when using the computer – you will feel more comfortable and you will work more efficiently. Also make room for the mouse if you are using one.
- Keep the computer cables tidy and out of the harm's way – don't leave them trailing across the floor where people might trip over them. If you have a problem with trailing cables, consider rearranging the room so that the computer is closer to the power points.

7.2 Care of the equipment

- Avoid dropping, knocking or shaking the equipment.
- Do not drink any tea, coffee or other liquids when using the computer – any spillage on the equipment (especially the keyboard) can damage the equipment.
- If you work in an area where power cuts or voltage fluctuations occur, consider buying a UPS and a voltage regulator. These will help protect your equipment *and* your data.
- Place a dust cover over the equipment when it is not in use (but remember to switch all equipment off first!)
- Avoid using or storing the computer in hot or cold areas. Also, keep the equipment out of direct sunlight.
- When connecting any peripheral equipment to the computer, do this carefully and do not force the plugs into the sockets. For absolute safety, switch everything off before carrying out the process.
- Some cleaning fluids can tarnish the surface of the computer and the computer monitor so make sure you use the correct type of fluid.

7.3 Moving the computer

- Unless the computer is a portable one, avoid moving the computer more than is absolutely necessary.
- Be careful when moving your computer – make sure that it is insulated from any vibrations or knocks by using the original packing material or carrying case.
- Place a blank floppy disk or the shipping insert in the floppy drive to protect it.

In Chapter 10 we will look at data security and ways that you can ensure data integrity.

8 EXERCISES

Exercises

1. Indicate whether the following statements are true or false:
   a. Computers contain microprocessors
   b. Instructions can be given to the computer using the keyboard or mouse
   c. The computer's memory is the amount of space contained on the hard disk
   d. A floppy disk often contains 100 times more space than a hard disk
   e. CD-ROM's have a large capacity for data and are used for literature databases
   f. The operating system can control the operation of all equipment attached to the system unit
   g. Operating systems are designed to work with different families of microprocessors
   h. In general, software written for an older version of an operating system will not work with newer versions of the operating system
   i. In general, software written for an older generation of a microprocessor will not work with newer generations of the microprocessor
   j. The larger a disk is in physical size, the more data it can hold
2. Fill in the missing words:
   a. The computer's _____ is its electronic working space
   b. The _____ is the principal microprocessor or "brain" which controls the other parts of the system
   c. A _____ is a small, hand-held plastic box which, when moved on a flat surface, controls the movement of a pointer on the screen
   d. Most programs and data are stored on the _____ _____which is located inside the system unit
   e. Compact disks from audio technology have been adapted to store large amounts of data in systems called _____
   f. If you wanted to communicate with another computer over the telephone line you would need to use a _____
   g. The _____ _____ is an area of the disk which is referenced when the computer starts up. It is also a target for a particular type of virus.
   h. Viruses can easily be spread between computers over a network or by using infected _____ _____
3. List the hardware that you would expect to find in a typical microcomputer set-up. State briefly what the function of each item is.

EXERCISES

4. What are the differences between hard disks and floppy disks? What other technologies are used to store and retrieve data?

5. List the different types of printers that are available. Which types are suitable for printing high quality computer graphics?

6. What are the possible consequences on a typical computer set-up of fluctuations in power? How could you avoid these consequences?

7. Data can be entered into the computer via the keyboard or with the aid of a mouse. What other ways are there of getting data into the computer?

8. List the different types of computer viruses and their features.

9. Some floppy disks will work on one computer but not on another. List possible reasons for this.

10. If you are installing a new computer in your office, what steps would you take to ensure the safety of people using the computer and the safety of the computer itself?

Chapter **8**

Database basics

In this chapter, we will be looking in detail at how you can build a computerized documentation system for defined groups of accession-specific data using database management software. We will also look briefly at how you can handle group data. When you have finished this chapter, you will be able to:

- Explain the terms database, file, table, record, field
- List the basic capabilities of database software
- Be aware of the differences between flat file and relational database management software
- State the main principles used in defining databases
- Describe the benefits of using of indexes
- Design file structures so that separate files can reference each other
- Link data files using chosen descriptors
- Describe the considerations to be taken when defining fields
- Produce a list of field definitions
- Outline approaches for the documentation of group data
- Explore how spreadsheets can be used with databases to perform certain tasks not possible with some database software alone

1 Introduction

Definition: COMPUTER DATABASE

A well-organised set of interrelated data held in one or more files which are capable of being managed by the same software. Files managed by different software are completely separate databases.

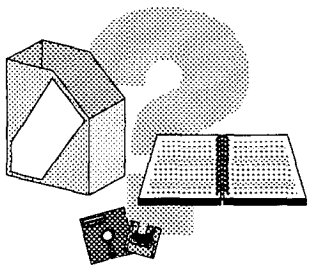
1.1 Importance of good organisation

The creation of files which have a clearly defined purpose and a logical structure is fundamental to good database design.

Different descriptors which are used in conjunction with each other in a given procedure, or related to each other in some way can normally be grouped together in a single database file. In many cases this is fairly obvious, for example you wouldn't have inventory data in a file used for germplasm evaluation.

Each file must have a clearly defined structure for storing the data. This will influence the ease with which data can be recorded, updated and retrieved. A badly designed structure will result in an inflexible, inefficient and problematic system.

1.2 Importance of software



The performance of a database depends on the capabilities and features of the management software used. Since capabilities can vary considerably between different software packages, it is vital to spend time studying available software to ensure it suits your needs. Remember that much of the commercially available database software can be tailored to meet the specific requirements of your genebank documentation system.

1.3 Files, records and fields

Your genebank analysis will have identified meaningful sets of descriptors which are practical in terms of recording data and practical in terms of retrieving information. The simplest example of such a set is a single page layout in a manual documentation system and a corresponding-file in a computerized system.

Consider the example of seed drying which was described in Chapter 4. You might define a descriptor list consisting of the following descriptors:

Seed Drying

Accession number
Batch reference
Final moisture content
Date of final moisture content determination
1000 seed weight

Having defined this descriptor list, you can design a page layout in a manual system, perhaps something like fig. 1.

Accession number	Batch reference	Date of final moisture content determination	1000 Seed weight (g)	Final moisture content (%)
1363	15-Oct-1990	17-Sep-1991	527	7
1427	08-Nov-1990	17-Sep-1991	692	6

Fig. 1. A manual form designed to document a descriptor list for seed drying

Each file in a database can be thought of in the same way as in fig. 1. In fact, frequently each file is called a *table.*

A *field* is the section of the file (or table) which always holds the same descriptor. Each column in fig. 2 can be thought of as a field, each descriptor occupies a field in the table.

Accession number	Batch reference	Date of final moisture content determination	1000 Seed weight (g)	Final moisture content (%)
1363	15-Oct-1990	17-Sep-1991	527	7
1427	08-Nov-1990	17-Sep-1991	692	6

Fig. 2. A field is indicated by the shaded area

A *record* is a set of different fields which are handled as a unit. A record holds different descriptors which relate to a single element. A record corresponds to a row in fig. 3.

Accession number	Batch reference	Date of final moisture content determination	1000 Seed weight (g)	Final moisture content (%)
1363	15-Oct-1990	17-Sep-1991	527	7
1427	08-Nov-1990	17-Sep-1991	692	6

Fig. 3. A record is indicated by the shaded area

The example in fig. 1 contains two records and five fields: there are records for accession numbers 1363 and 1427. There are fields for accession number, batch reference, date of final moisture content determination, 1000 seed weight and final moisture content.

When designing a manual form, you have to decide which descriptors should appear on the form and how wide each column should be so that it can comfortably hold the expected data. With computerized databases, it is also necessary to specify how data should be handled in different fields, but as we will see later in this chapter, there are many more characteristics to specify.

Often the file design in a computerized documentation system resembles the corresponding form of the manual documentation system. Usually the same set of descriptors are used. Your analysis defined the set to be practical in terms of recording data and practical in terms of retrieving information. This will be true-for both manual and computerized systems.

1.4 Features of database management software

Most database management software allows you to perform the following basic activities:

- Enter new data
- Modify or delete data
- Search and retrieve data for reports
- Sort data
- Import and export data
- Modify the structure of a file in response to changing information needs

Database software vary according to:

1. How flexibly the software can perform these activities
2. How easy the software is to use
3. Additional features of the software such as file security and shared access to files

You'll come across two main categories of database management software: *flat file managers* and *relational database managers*, which will be looked at in detail in the following two sections.

1.5 Flat file managers

Flat file managers are the more simple type of database manager and generally the easiest to understand and use. Each database consists of a single file. You can perform the basic activities listed above on each file in a flexible way, but only usually on one file at a time. In many cases this is quite adequate, but you will meet the limitations of a flat file system when you need to work with data from different files at the same time.

With many currently available flat file managers, there are ways around this difficulty. It is usually possible to set up your system so that different files can make reference to each other and in effect, can be worked with at the same time. How files can be linked should be looked at right at the start, so that allowances can be made when you design your basic database structure. Study your database software manual carefully for instructions on how to do this.

If you intend to use flat file database management software and wish to set up links between files, ensure first that-your software is able to do this.

1.6 Relational database managers

Relational database managers differ from flat file managers in that they are designed specifically to work on more than one file at the same time. You can perform the basic activities listed above on more than one file at a time. This is achieved by linking two or more separate files through a field which is shared between files. By linking the files together, a *relationship* is produced between the files. The shared field is only stored once, it is not duplicated in each file, and acts as the link between the separate files. The linked files can then be worked on simultaneously. The technical term for the process of linking files in this way is *normalization.*

The theory of relational database managers is fairly complex. The above description is a simplified view, but should be sufficient for our current needs.

1.7 Linking files together

When designing your genebank documentation system, you are likely to want to link files together so that you can work with data from different files at the same time. You will do this in different ways depending on the type of software you have chosen (a flat file manager or a relational database manager) for the reasons outlined above.

You might have the following separate files for a particular crop:

Registration	Passport	Characterisation
Accession number	Accession number	Accession number
Scientific name	Collector's number	Batch reference
Crop name	Collecting institute	Sowing date
Depositor number	Date of collection	Soil type
Acquisition date	Country of collection	Soil *p*H
(other fields)	(other fields)	(other fields)

Suppose you were producing a report on a characterisation trial for a series of accessions and you wanted to include some registration and passport data as well. How would you do this? Rather than searching the three files separately and producing three separate reports which would then have to be combined into a single report, think about how you could produce a single report by linking the files together.

In a genebank documentation system, a field which is often used to link files together is the accession number:

Registration	Passport	Characterisation
Accession number ◆	Accession number ◆	Accession number ◆
Scientific name	Collector's number	Batch reference
Crop name	Collecting institute	Sowing date
Depositor number	Date of collection	Soil type
Acquisition date	Country of collection	Soil *p*H
(other fields)	(other fields)	(other fields)

Here, a record in one file can be linked to one or more records in other files which have the same accession number. Of course the accession number isn't always used as the linking field, it could be any other field which is common to different files. You could even use more than one field to establish links.

Another important feature of linked files is that you can often use one file to modify data in another file. For example, you could have a file which details the distribution of seed from the seed store and another file which details the inventory of the seed store:

Movements	Inventory
Accession number ◆	Accession number ◆
Batch reference ◆	Batch reference ◆
Date of supply	Location
Amount of seed sent	**Total weight of seed**
Recipient	(other fields)
(other fields)	

To uniquely identify a batch, it is necessary to use both the accession number *and* the batch reference to link the two files together. Then you could instruct the field "amount of seed sent" to update the inventory file so that it indicates the correct weight of seed left in the store. Consult your database software manual for instructions on how to do this.

The ability to link files is a very powerful resource

The ability to link files is therefore a very powerful resource. For example, it can be very useful for documenting accession-specific data and for retrieving information stored in different files, which can then be manipulated into the form you need. Let's now look in greater detail how you can organise and use linked files in genebank documentation.

2 Organising your files

2.1 Identifying fields

When files are being linked together to produce reports or to update data, the software has to select the right records to work with. How does it do this?

Look at a registration file which contains one record per accession, an inventory file which can contain one or more records for each accession (because the seed store might contain one or more batches) and an evaluation file which could contain several records for each accession (representing evaluation trials performed on different batches at different times). How could these files be linked?

Often you will want to identify and work with one individual record. But sometimes you may want to work with a group of records that, for example, have the same accession number. How can you instruct the software to select the correct record or set of records? Is more than one field needed to distinguish one individual record or group of records from another?

It is important to clarify which field or combination of fields are needed to identify the records you wish to work with. Such fields are called *identifying fields* and are used to link different files together. How links are set up will depend on whether the identification of a specific record is required or whether you wish to work with a group of records. Examples of identifying fields are listed in table 1.

Table 1. Examples of possible identifying fields for different files

TYPE OF FILE	NO. OF RECORDS/ACCESSION	IDENTIFYING FIELDS
Registration	1	Accession number
Inventory	1<	Accession number Batch reference
Characterisation/evaluation	1<<	Accession number Batch reference Date of test
Inventory/movements	1<	Accession number Batch reference Date of distribution

2.2 Guidelines for design of data file structure

We should now look again at the meaningful sets of descriptors identified in your genebank analysis. Their structure should be used as a basis for building your documentation system, but they may need some restructuring to take advantage of the power of linking files.

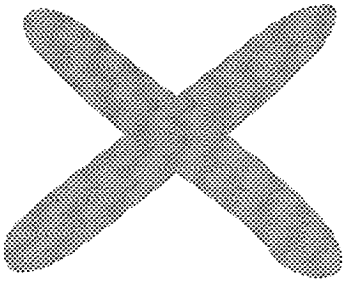
Can each set of descriptors be used for a separate file in the database? The answer is "yes" if the following principles are followed:

1. Avoid duplicating unlinked fields in separate files

We have discussed how files can be linked through common fields. However, you must think carefully about which fields you are using to link files and avoid duplicating data unnecessarily.

Unlinked fields which are duplicated in two or more files contain redundant data, which waste valuable space on your disk. Also, if such fields need updating at any time, you must remember to update each in turn, otherwise inconsistencies will start creeping into your database. It can also be difficult to remember where the duplicated fields are located. Avoid these inconsistencies right from the start by not duplicating unlinked fields in separate files.

Let's look at a genebank example. Consider the following files for registration, passport and characterisation of a *Phaseolus* species. There are three fields which are duplicated across the files, namely: accession number, scientific name and crop name.



Registration	Passport	Characterisation
Accession number ◆	Accession number ◆	Accession number ◆
Scientific name ◆	Scientific name ◆	Batch reference
Crop name ◆	Crop name ◆	Scientific name ◆
Pedigree/cultivar	Collector's number	Crop name ◆
Depositor number	Collecting institute	Sowing date
Acquisition date	Date of collection	Harvest date
Type of material	Country of collection	Evaluation environment
(other fields)	(other fields)	Soil type
		(other fields)

You need the accession number in each file as this is the field which will be used as the identifying field. However, the scientific name and crop name are unnecessary duplications as they will always be the same for a particular accession. Imagine what would happen if the name of the *Phaseolus* species changed. The scientific name would have to be amended in three of your files. If you didn't change the scientific name in all three files, there would be confusion later on over which is the correct scientific name.

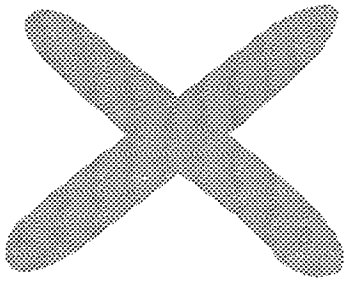
Changes like this will be a much easier for you to record if the scientific name is only stored in one file. The same is true for the crop name. In this case, the most appropriate place to store these names is in the registration file, particularly if this file is a general registration file for all genebank accessions.

2. Fields which have a logical relation to each other and to the identifying field, should be grouped together in one file

Suppose you had a file called "movements" listing movements from the seed store which had the following fields:

Movements

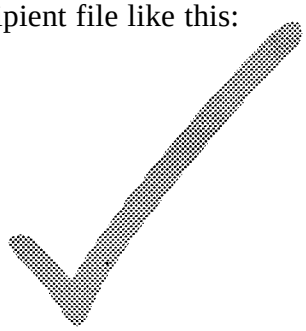
Accession number
Batch reference
Date of supply
Amount of seed sent
Recipient name
Institute name
Address line 1
Address line 2
(other fields)



"Accession number" and "Batch reference" are the identifying fields since they identify precisely which seeds were sent out. However, the fields "institute name, address line 1, address line 2" are information

about the "recipient name", not the accession number. If the "recipient" requested seeds at a later date, the entire address would need to be entered again in a different record. It would be much better to store information about the recipients in a separate recipient file like this:

Movements	**Recipient**
Accession number	**Recipient code** ◆
Batch reference	Recipient name
Date of supply	Institute name
Amount of seed sent	Address line 1
Recipient code ◆	Address line 2
(other fields)	(other fields)



Each recipient could be given a recipient code which uniquely identifies them. (A bank account-number identifies a particular person's account in a similar way). When seeds are sent out, the recipient code is stored in the movements file. The recipient file, where the address is stored, is linked to the movements file through the recipient code.

Let's study a more complex example. In Chapter 6 we looked at the structure of a management file for a manual documentation system which contained summary data on inventory, seed viability and seed drying. The descriptor list looked like the list in table 2.

Table 2. **Example descriptor list for a management file in a manual documentation system. Shaded areas indicate those descriptors that would be more efficiently stored in a separate, but linked, database file**

DESCRIPTOR
Accession number
Scientific name
Batch reference
Collection type
Location in store
Total weight of seed
1000 seed weight
Minimum amount of seed allowed
Moisture content
Viability (%)
Date of viability test
Date of next viability test
Date of supply (distribution)
Amount of seed sent (distribution)
Recipient code (distribution)
(Other descriptors)

By linking together the files of computerized databases, you **CAN** look in several places at the same time

The reason we suggested this type of file for a *manual* system was for practical reasons – you wouldn't have to look in three separate places to retrieve information. But with computerized databases, by linking files together you *can* look in several places at the same time. One large file like the one above would be clumsy to work with, while several smaller files linked by common fields would be more efficient at data management.

The shaded fields in table 2 are not directly related to either of the identifying fields (accession number and batch reference). They are concerned with viability and seed distribution. A better structure would be to use three files instead of one. You could design your files as follows:

Inventory	Seed viability	Movements
Accession number ◆	Accession number ◆	Accession number ◆
Batch reference ◆	Batch reference ◆	Batch reference ◆
Location in store	Viability	Date of supply
Total weight of seed	Date of viability test	Amount of seed sent
1000 seed weight	Date of next viability test	Recipient code
Minimum weight of seed allowed	(other fields)	(other fields)
(other fields)		

The fields in each file are dependent on only the identifying fields (accession number and batch reference), which are used to form the link between the files.

2.3 Sorting

In a manual documentation system, we have seen that records are stored according to the following rules:

1. **Records must be stored in an order that will help subsequent information retrieval**

2. **For accession-specific data, the most useful order for storage is in order of the accession number**

With a computerized system, when new records are allocated an accession number, they are simply added to the end of the file. This means that records (identified by accession number) can theoretically appear in any order as illustrated in the characterisation file in fig. 4.

Accession number	Batch reference	Sowing date	Soil type	Other fields
1515	08-Nov-1987	28-Jun-1988	3	
573	18-Jul-1988	07-Sep-1990	3	
1369	08-Nov-1987	07-Sep-1990	3	

Fig. 4. Records of a computerized characterisation file appearing in random order (not sorted by accession number)

This random ordering is not a problem for a computerized system in the way it would be for a manual system. Before information is retrieved, the computer can be instructed to *sort* the records into a specified order (for example, accession number order). So, the two rules ex-

plained above which must be applied to manual systems are unnecessary for computerized systems.

Moreover, a computerized system can be instructed to sort on other fields apart from the accession number. This would allow you, for example, to sort your records in order of batch reference which could be useful if you needed to know which accessions were in need of regeneration. It is even possible to use more than one field to sort by. You could, for example, sort by both accession number and date of last viability test (see fig. 5).

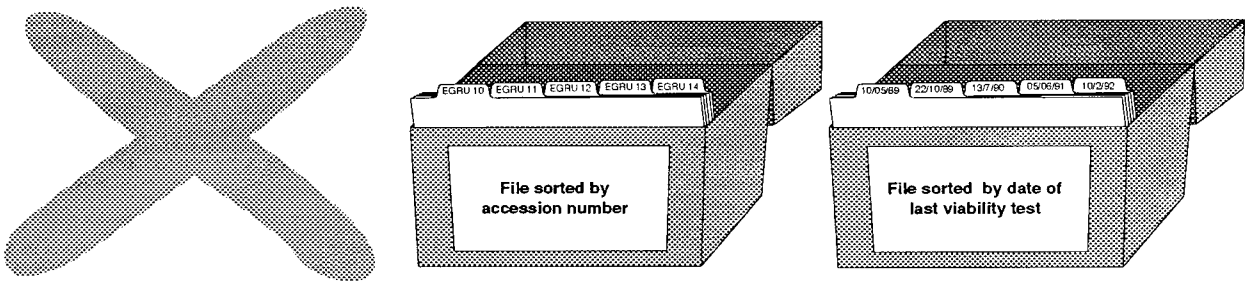


Fig. 5. Sorting a file by both "accession number" and "date of last viability test" is possible with a computerized, but not with a manual documentation system

Records can be sorted in different ways. Commonly they are sorted in alphabetical, numeric (ascending or descending) or date order. There are other ways of sorting, but as all database software differs in the way it performs the sorting process, you should consult your manual at this stage for more information on the capabilities of your system.

2.4 The use of indexes

An *index* is usually a separate file which stores information on the location of the specific records in a database file. It is rather like an index in a book which lists the page numbers of the different subjects. Indexing is used to help you locate specific information in your database.

Indexing is carried out on selected fields, across all the records in a database file. If you were to index the accession number field, an index file would be created containing a reference to the location of each record (identified by accession number) in your main database file. This reference to the location of a record is often referred to as an *address.*

Indexes are used in subtly different ways by different database managers, so again, it is worth spending some time reading your manual to understand how indexes are used in your software. Often indexing is

used to speed up searching. However, some database managers will not search on a field *unless* it has been indexed.

With most software, you can choose to index as many fields as you wish. Indexed fields will usually be fields which are accessed frequently. So, identifying fields, discussed previously in section 2.1, should always be indexed as they are worked with very regularly.

3 Relating your files

After looking at the importance of good file organisation and how this can be achieved, we are now in a position to see how the files can relate to each other.

From your genebank analysis, you have probably identified a number of files including:

- A single registration file which can be used for *all* genebank accessions
- A number of passport files – one for each crop
- A number of characterisation/evaluation files – one or more for each crop
- A single inventory file which can be used for all genebank accessions
- A number of files concerned with seed testing
- A file detailing movements from the seed store (where appropriate)
- A file detailing recipient addresses (where appropriate)
- Group data (to be discussed in section 5 of this chapter)

For the first four files listed above, we can visualise the relationship for three crops (A, B, C) each having three characterisation or evaluation files (1,2,3) as in fig. 6.

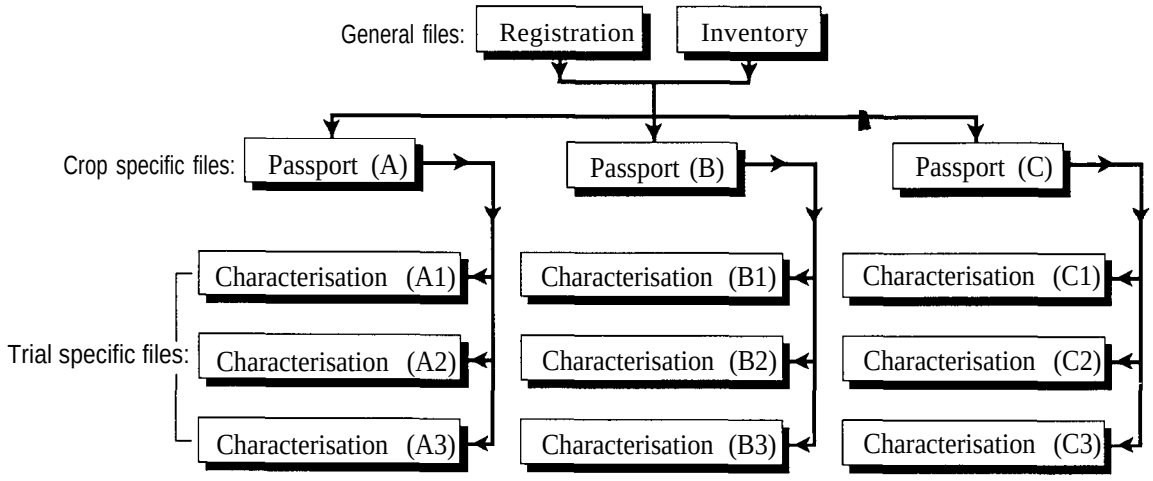


Fig. 6. The relationship between registration, inventory, passport and characterisation/evaluation files for three crops: A, B, C

How can you meaningfully relate these files? Which fields should be used to form the links? If it is not already clear, table 3 should help.

Table 3. Identifying fields which should be used for linking different data files

IDENTIFYING FIELD: / FILES:	ACCESSION NUMBER	BATCH REFERENCE	DATE OF TEST
Registration	✔		
Passport	✔		
Characterisation	✔	✔	✔
Inventory	✔	✔	
Seed handling	✔	✔	✔
Evaluation	✔	✔	✔

In order to retrieve information in a useful form, the relation between files must be **MEANINGFUL**

The concept behind relating files together is that in order to retrieve information in a useful form, the relation must be *meaningful.*

For example, there is no point in relating a *Phaseolus* characterisation file with a *Musa* characterisation file using the "Batch reference" as the identifying field, even though it could theoretically be done. You would retrieve a mixture of data from unrelated accessions (which coincidentally have the same batch reference) which would be neither useful nor meaningful. So, think carefully about how you establish relations between files!

Look back at table 1 on page 160 for examples of which fields can be used meaningfully as identifying fields for different types of files.

4 Defining fields

Once you have decided how to organise your files, you can start to specify the characteristics of all fields contained in each file. It is useful to produce a *record definition,* which details the structure of the record and the characteristics of each field, before you begin to build the main database file.

When you define fields you should give careful consideration to the following:

- The name of each field
- The type of data a field can hold. For example, numeric, character, date, logical (yes/no or true/false), etc.

- Whether a field will be used as a link to another database file

The decisions you take here will affect the way the database operates, so we will look closely at each of these characteristics in the following sections.

4.1 Field names

The name of a field is important as it is used when you link files together, build reports and design entry forms.

A field name should be:

- Unique – it cannot be confused with another field name (by either the software or the operator)
- Descriptive – it describes the field contents
- Simple – so you can remember it easily

Database software often places a restriction on the length of the field name allowed. For example, some database managers only allow 10 characters for the field name, so "Current moisture content" would not be acceptable. In this case, you would have to use an abbreviation for the field name.

Also, some database software will not accept spaces between words in the field name. If this is so and you wish to separate words, you can get round this by using an underscore (_) between words.

4.1.1 The name of a field must be unique

NEVER duplicate field names between files unless they refer to the same descriptor

Never duplicate field names between files unless they refer to the same descriptor. This rule must not be broken.

For example, if the fields "accession number" and "account number" were both shortened to "act_no", neither you nor your software would be able to distinguish between them when searching, retrieving information, producing reports, linking fields together, etc. It would result in serious confusion.

4.1.2 Choose a field name which is descriptive

Avoid using codes or highly abbreviated names

You should avoid using codes or highly abbreviated names as they can make it difficult for you and other people to identify a field. For example, don't use "X", "F4", or "A_D" for "Acquisition date". The first two bear no resemblance to "Acquisition date" and the third is so abbreviated that it isn't clear which field it refers to. "Acq_date" would be more descriptive and much easier to recognise.

4.7.3 Be consistent in naming fields

Use the same abbreviation each time

If you have to abbreviate a word in several different field names, use the same abbreviation each time. It will make it easier for you to recognise the field.

For example, field names containing the word "date" are common in genebank documentation. Try to adopt a convention of handling the word in the same way each time. Do not use "acq_date", "sowing_dat", "regener_da", it would be much better to use "acq_date", "sow_date" and "regen_date".

Make sure that a given descriptor has the same field name in different files

Make sure that a given descriptor has the same field name in different files. This will allow you to link files using these fields. Report writing will also be much more straightforward: This may be obvious for descriptors which are commonly used as identifying fields (such as accession number, batch, date of test), but it is equally important for crop-specific passport fields such as collector's number, collecting institute, collecting site, country of collection, latitude, longitude, altitude, etc.

4.2 The type of field

Each field can be assigned different characteristics which considerably change the way in which it handles the data it holds. The most important characteristic is the *type of field.* Different database managers use slightly different types of fields, but a typical list is given below:

- *Character* – anything that can be typed at the keyboard (including digits)
- *Numeric* – for digits only
- *Scientific notation* – for very large or very small numbers (e.g. exponential numbers)
- *Date* – handles the date (year, month, day)
- *Logical* – for true/false (or yes/no) data

Certain fields will only accept data in a specific format. For example, a date field might be set up to only accept a date in the form DD/MM/YY and not MM/DD/YY. Logical fields will often only accept a specific set of responses.

It's usually a straightforward process to decide what type of field to use for a given descriptor, often only one type is possible. However, there are cases where you can choose so you need to know the relative merits of each choice. Some considerations are as follows.

4.2.1 Identical fields should have identical specifications

This is very important. Fields with the same name should have identical specifications, especially if you want to use the field to relate files. You wouldn't define the "accession number" in one file as a character field and as a numeric field in another; if you did, you might not be able to relate files together.

4.2.2 Use numeric fields for digits only

You should examine carefully your descriptor states to find out whether there are cases where you may have to enter a character rather than a number.

For example, this could occur when you have defined an ordinal scale of 0-9 for a descriptor, zero being used for "absent" or "not-expressed" but you have not performed the test. Then you would want to indicate "unknown" or "not performed". If you leave the field blank, some software will automatically store a zero which means "not expressed". If you try to store "?" for "unknown", the software will not accept the response as you have defined the field to only accept digits. In such cases, it would be better to define the field as a character field.

4.2.3 Do not use character fields for purely numeric data

Character fields cannot usually be used directly in mathematical calculations. If digits are held in a character field, you may find that the contents of the field need to be converted before your software can perform calculations. Some database managers even require you to specify whether you want to be able to perform calculations on a numeric field. So, if you wish to perform calculations on a field, make sure that the field is defined to enable you to do this.

If the contents of a field will only ever be digits, the field should be defined as numeric

When character fields are used to store numeric data, the mistake of someone entering a character in place of a digit, or a space where one should not be, cannot be avoided. For example "1OO6" instead of "1006" or "9 11" instead of "911". If the contents of a field will only ever be digits, the field should be defined as numeric.

4.2.4 Use "memo" or "comments" fields sparingly

Some database management software will allow you to define different types of text fields which have different characteristics.

Sometimes you might want to store a large amount of text, or entries that vary in length and cannot be standardised with "acceptable values". Some software will allow you to define a specific field for such special cases. This type of field can have various names, but is often referred to as a *memo, comments* or *remarks* field.

There is a temptation to use this type of field in each file to handle miscellaneous data. They should, however, be used only when absolutely essential. Problems can arise when you want to locate data in such a field. Some software requires you to search this field separately in each record – which can mean a lot of extra work. Some software will not allow you to search this type of field at all.

It is therefore advisable to restrict the use of these fields. Perhaps just include one in the registration file which you would use for all your miscellaneous data; that way you'd only have to look in one place. It would be something like having an information sheet for each accession in a manual documentation system (see Chapter 6).

Table 4 gives examples of field types that can be assigned to different descriptors.

Table 4. Possible field types for a range of descriptors

FIELD TYPE: / DESCRIPTOR:	CHARACTER	NUMERIC	DATE	LOGICAL	MEMO/COMMENTS
Accession number	✔				
Donor identification number	✔				
Acquisition date			✔		
Photograph taken				✔	
Collector's notes					✔
Soil *p*H		✔			
Monthly rainfall		✔			
Flower colour	✔	✔			
Local name	✔				

4.3 The size of the field

Before defining fields, you should check if your software requires you to specify field sizes

Some database managers, require you to specify from the start how much data (i.e. how many characters) a field can hold. This is often called the *field width*. With other database managers, fields can 'grow to fit' the amount of data input. Before you start defining fields, you should check if your software requires you to specify field sizes. If it does, you need to give careful thought at this stage to the maximum amount of data you are likely to want to store in each field. However, try not to overestimate the size as most software will allocate this amount of space to each record, irrespective of whether any data is stored, which wastes disk space.

For example, if the highest genebank number is currently "1324" you should choose a field width of "5" rather than "4" particularly if your genebank is expanding.

Usually, the width of numeric fields includes the number of decimal places, the decimal point and the sign (+/-) used. So, "2312" has a field width of 4, "109.2" has a field width of 5, "-0.712" has a field width of 6.

An important point that was raised earlier is that fields with the same name should have identical specifications. If your software requires you to specify the field width, this must also be identical for fields with the same name. If not, data can be lost when it is transferred from one field to another, if the receiving field has a smaller field width.

If after working with your system, you find that the fields you have defined are too narrow, they can usually be adjusted at a later date. Check with your software manual how to do this.

4.4 Scientific name

Store the scientific name for each accession in one file only

We discussed earlier how it is laborious and problematic to keep a field up to date if it occurs unlinked in more than one file. This is particularly true of scientific names which are prone to change. Therefore it is advisable to store the scientific name for each accession in one file only. The registration file is the most appropriate place for this. Then if you need to change the scientific name you only have to go to one place.

Another consideration is how to define the field for the scientific name. Some scientific names are very short (e.g. *Zea mays)* and others are very long (e.g. *Macroptilium longepedunculatum* or *Brassica oleracea* var. *gemmifera).* If you are required to specify the field width, how wide should the field be? Should you use abbreviations? Should the scientific name be divided into several fields?

Abbreviating the scientific name will save space, but could easily become inconsistent and confusing and so should be avoided. For example, with *M.longepedunculatum,* does *M.* stand for *Macroptilium, Mangifera, Manihot, Medicago, Mentha, Mimosa, Musa...?*

Avoid confusion – **DO NOT** use abbreviations

Remember, your first concern should be to have reliable data. So, avoid confusion – *do not* use abbreviations.

For maximum flexibility, consider using the separate fields of genus, species, sub-species and sub-species rank as illustrated in table 5.

Table 5. Separate fields for genus, species, sub-species rank and sub-species

	GENUS	SPECIES	SUB-SPECIES RANK	SUB-SPECIES
Zea mays	*Zea*	*mays*		
Macroptilium longepedunculatum	*Macroptilium*	*longepedunculatum*		
Medicago* sativa subsp. *falcata	*Medicago*	*sativa*	subsp.	*falcata*
Brassica oleracea* var. *gemmifera	*Brassica*	*oleracea*	var.	*gemmifera*

If you decide to follow this convention, the following field widths should accommodate all the scientific names you will encounter:

genus:	24 characters
species:	26 characters
subspecies:	26 characters
subspecies rank:	8 characters

Make additional fields for the authority for the scientific name if you decide you need to record this information.

Choose a format that is convenient to use and avoids any ambiguity. Ask yourself how you want the scientific name to appear in the different reports – is it always in the same way? Do you always print the sub-species name or do you sometimes leave it out? Do you print abbreviated genus names or is it necessary to print them in full?

4.5 Practical hints

Consistency in defining fields is essential. One way to ensure this is to make a register of field definitions for each file. For each field you can record:

- Full descriptor name – e.g. “Accession number”
- Field name (usually an abbreviation) – e.g. “act_no”
- Field type – e.g. character, numeric, logical, date, etc.
- Field description – an explanation of how the field should be used, including what type of data should be entered and in what form
- Data validation rules – any rules for data validation which apply to the field
- Index – is the field indexed?
- Field width (where appropriate)
- File name – in which database file the field occurs

Most database management software can automatically generate this list of field specifications for each file, which is referred to as a *data dictionary*. If your software cannot do this, you could do it on paper. Design a form with columns for each of the above descriptions and simply fill in the field definitions for each descriptor. A data dictionary for a passport file might look like table 6.

Produce a data dictionary for each file in the database

You should produce a list like this for each file in the database. You’ll consult these lists every time you define a new field so that consistency is ensured across your files. If you already have some experience with database software you could even use your file manager to maintain a central register of field definitions!

Table 6. A possible data dictionary for a passport file

DESCRIPTOR NAME	FIELD NAME	FIELD TYPE	DATA VALIDATION	FIELD DESCRIPTION	FIELD WIDTH	INDEX
Accession number	acc_no	Numeric	Between 1 and 999, 999	Unique identifier assigned to the accession when it enters the collection	6	✔
Date of collection	coll_date	Date		Date of collection of original sample	(software defined)	
Collection source	coll_source	Character	1-8	Collection source using the following table: l=wild habitat 2=farm land 3=farm store 4=backyard 5=village market 6=commercial market 7=Institute 8=other	1	
Photograph taken	photograph	Logical	0,+	Was a photograph taken of the accession or environment at collection?	1	
Weight of seed collected (g)	coll_wt	Numeric	Between 0 and 99, 999		5	
(other fields)						

5 Documentation of group data

Group data are data concerned with groups of accessions rather than individual accessions. The data might refer to a particular species or crop, for example; viability test methods and regeneration procedures. The group data might even refer to several genera and species such as information about the flora of an ecogeographical region or information from ethnobotanical studies. This kind of data is likely to be reproduced in publications such as journals, books, and conference proceedings.

How can you use a computer to manage these data? As with any data that you handle, you have to determine your information needs and then form meaningfully related sets of data (descriptor lists), that will be practical in terms of recording data and practical in terms of retrieving information.

As these sources of data are so diverse and the subject area is so enormous, your information needs will be many and varied and the number of descriptor lists for group data will be potentially very large. It's unrealistic to even think of computerizing all the relevant data from the published material. It's more practical to keep a reference to the source of the publication which should be maintained separately from your documentation database. This will be a literature database based on a "descriptor list" for literature references. You'll design the descriptor list based on your information needs and the kind of references you want to computerize.

Keep a reference to the source of the publication

Literature databases are often maintained using database software developed specifically for handling text – so-called *text management systems.* If you have access to such a system, it is worth thinking about using it to construct your literature database. If you don't have access to such a system, look at how your existing database software can be used and adapted to do the job. If you use your existing software, remember that your literature database should be a separate database file.

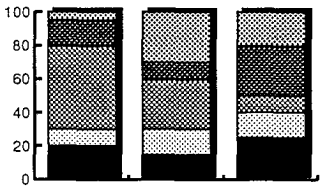
Your literature database should be a separate database file

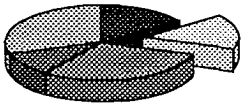
Text management systems operate in much the same way as other database software but two powerful features are worth highlighting. Firstly, they have greater flexibility in the way that they perform searches, as you don't always have to tell the software which fields to search, it can be set up to look in every field. This is very useful if you need information but don't know where to look. The second feature concerns the way the software stores the data: fields "grow to fit" the amount of data entered and "empty" fields do not take up space on the disk. Literature databases commonly contain variable amounts of data for each record so this useful feature can result in enormous disk space savings.

6 Use of spreadsheets

	A	B	C	D	E
1	20/(100-x)				
2					
3					
4					
5					

Fig. 7. A typical spreadsheet as it appears on screen (with equation in cell Al)





Spreadsheets (see fig. 7) are powerful tools which are commonly used in both scientific and business applications to perform a whole variety of statistical and mathematical calculations. They can be used with data held in a database to perform more complex calculations than are possible with the database alone.

For example, you could export seed viability and storage data from your database to a spreadsheet, to calculate the effects of-varying the storage parameters on seed viability.

Spreadsheets can also be used to print reports or generate graphs using the data they hold. Therefore they provide a route to producing a graphical representation of the data stored in your database.

DO NOT use a spreadsheet as a database manager

Despite the obvious capabilities of spreadsheets, you should *not* use a spreadsheet as a database manager. Spreadsheets are specifically designed to be used for complex calculations, not for flexible management of documented data. With spreadsheets, it's usually not possible to relate different files, the search/report facilities of spreadsheets are limited and sometimes non-existent (which can make information retrieval very difficult), the modification of stored data can also be time consuming – especially when working with large data files.

If you want to store large amounts of data in several files, these limitations will cause you many problems. The use of spreadsheets should be restricted to the tasks they were designed for – performing complex calculations. For efficient documentation of your genebank data, you should stick with software designed specifically for this task – database management software.

7 Where to go from here

There are a number of steps involved in designing a database with linked files. These have been covered in this chapter and are summarised in fig. 8.

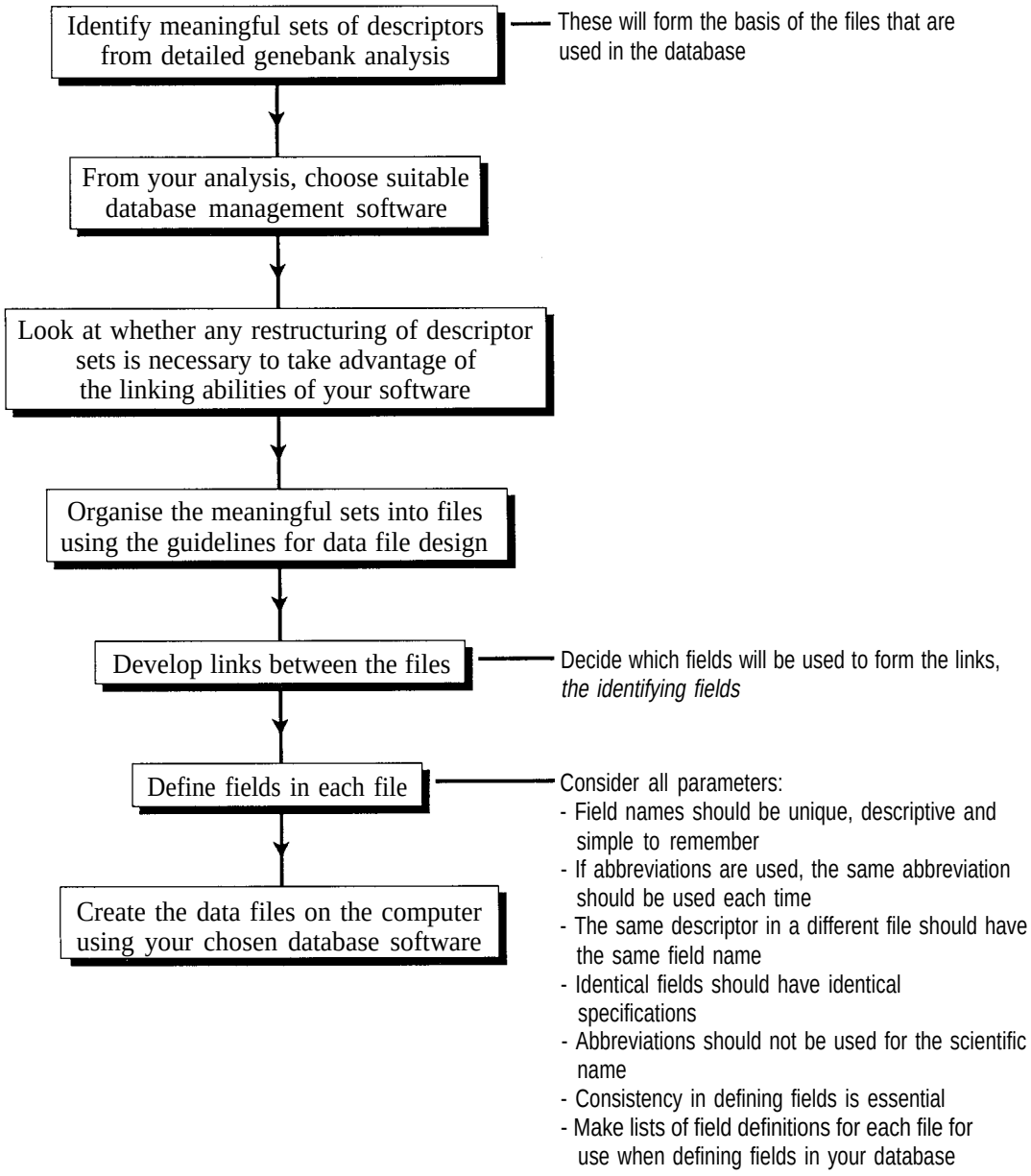


Fig. 8. The steps involved in designing a database with linked files

EXERCISES

8 Exercises

1. Indicate whether the following statements are true or false:
   a. A field in a file is sometimes called a table
   b. A record is a set of fields which are handled as a unit
   c. A computer database always contains several files
   d. Files can be linked together with any common field
   e. A registration file should contain only one record per accession number
   f. Fields of the same type (e.g. numeric, date) should have the same name
   g. It is important to avoid duplicating identifying fields in separate files in case of changes in nomenclature
   h. Indexing is often used to speed up searching
   i. Sorting is only allowed on identifying fields
   j. The batch reference is used as an identifying field for all accession specific data
2. Fill in the missing words:
   a. A _____ is the section of a file or table which always holds the same descriptor
   b. _____ _____ are used to select one or more records for the purposes of linking files together
   c. An _____ stores information on the location of specific records in a database file
   d. A _____ field can be set up to accept anything that is typed in at the keyboard including digits
   e. A_____ field accepts "yes", "no" data
3. What is the difference between flat file managers and relational database managers?
4. Identify situations where a set of descriptors identified in the genebank analysis *cannot* be used as a separate file in a database.
5. Illustrate how registration, passport, characterisation and inventory files can be related together in a database system, indicating which fields can be used to link the different files.
6. A register of field definitions is useful in ensuring consistency in defining fields. What details should such a register contain and why?

EXERCISES

7. Suggest field names for the following descriptors when your software only allows 10 characters for the field name:
   a. Accession number
   b. Acquisition date
   c. Altitude
   d. Collecting institute
   e. Collector's name
   f. Collector's number
   g. Collecting date
   h. Number of plants sampled
   i. Local/vernacular name
   j. Stem pigmentation
   k. Evaluation environment
   l. Location in store
   m. Frost damage susceptibility
   n. Name of person in charge of characterisation and preliminary evaluation
   o. Soil *p*H
   p. Pollination method
   q. Date of next test
   r. Field plot number
   s. Moisture content at harvest
   t. Moisture content at storage (initial)
   u. Duplication at other location

8. Spreadsheet software should not be used in place of database management software. Why is this? How could you use a spreadsheet in conjunction with your database management software?

Building the system

Chapter 9 studies the process of building a computerized documentation system for your genebank.

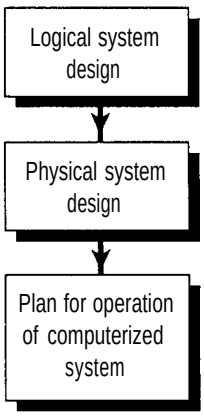
At the end of this chapter, we will have examined the following points:

- The stages of building a menu driven computerized documentation system
- Design considerations for screen forms, reports and menus
- Printing reports
- Implementing the data management features of your software
- Incorporating your screen forms and report formats into the routine procedures
- Development and organisation of menus

1 Introduction

In your genebank analysis, you will have studied:

- Data generation and the use of data in procedures
- The use of flow charts to clarify procedures
- The needs and priorities for documenting genebank data
- Information requirements (e.g. specific reports)
- The possibility of documentation being an integral part of genebank procedures



Studying these points provides a detailed list of specifications which should be used as the foundation for designing your computerized documentation system. This is sometimes called a *logical system design.* It defines how the documentation system should operate regardless of the hardware and software you are using.

From your logical system design, you can start to produce your *physical system design.* A physical system design details how the documentation system will operate with your chosen hardware and software. It also takes into consideration the people who will be

using the system. The end result of a physical system design is a detailed plan of how the computerized system will actually operate. It is this physical system design we'll look more closely at in this chapter.

The first step is to determine whether the software and hardware available to you can support your logical system design. There may be certain operational limitations. For example, the software may not be capable of performing the tasks you have specified, or the person selected to input data may not have access to a computer running the system. In such cases, you will need to make alterations to your logical system design to account for these complications. You might need to rethink your choice of software, or look at ways of reaching the same end result, using the same software in a different way. You might think of moving the person selected to input data so that he or she has access to a suitable computer, moving the computer nearer to the person assigned to input data or assigning the job of data input to another person.

1.1 Stages in building the system

By now, you will have defined the data files for your documentation system. The next step is to begin building the system, modelled on these files, using your chosen software. We must also look at how we can use the software to maintain these files and retrieve information from the system.

The different stages in building the system are illustrated in fig. 1.

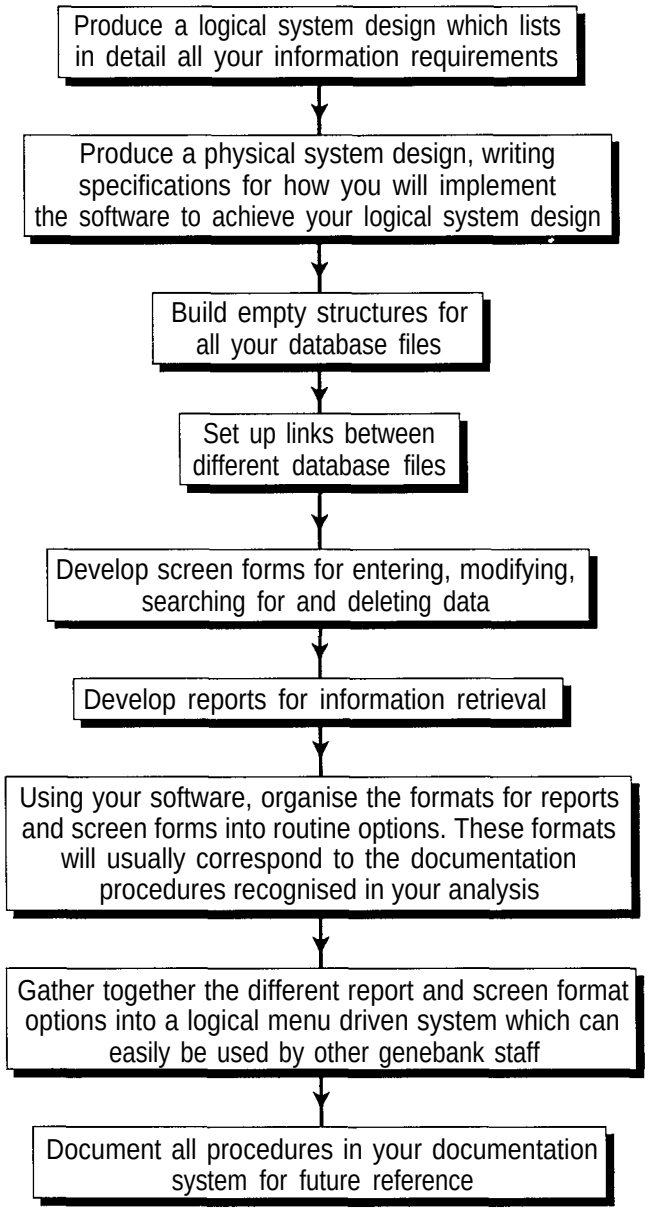


Fig. 1. The different stages in building a documentation system

In previous chapters, we have studied in detail the first four steps illustrated in fig. 1. In the following sections of this chapter, we will look more closely at the later steps of building the documentation system.

2 Developing screen forms

Recording and retrieving good quality, accurate and reliable data is fundamental to the success of your documentation system. If data recording is difficult or if there are no systems to check the quality of data entered, mistakes will inevitably be made and the information retrieved will be of doubtful quality. Correcting mistakes at a later date wastes valuable time, so it is vital to ensure data are entered correctly at the start.

Screen forms act as interfaces between the operator and the data files. They are also sometimes referred to as *views.* They are what you see on the screen when entering, modifying, searching for or deleting data; they are the computer equivalent of manual forms. Screen forms will help to ensure accurate and reliable data entry.

When designing screen forms, you should aim to:

1. Prevent errors getting into the system
2. Make data entry and modification "user-friendly"

As we'll see in the following sections, screen forms are actually very powerful and indispensable interfaces. A typical screen form is illustrated in fig. 2.

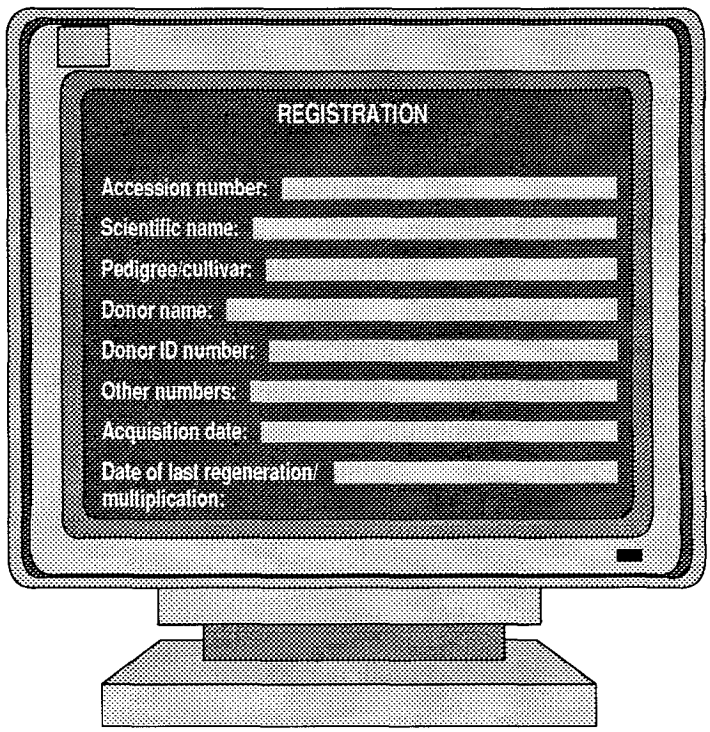


Fig. 2. A typical screen form

2.1 The need to design your own screen forms

Basic screen forms for data entry and modification are sometimes supplied with database management software. However, generally, it is much better to design your own forms.

Most database managers automatically generate a basic data entry screen which displays all the fields you have defined in the order they were defined. The field name allocated to each field usually appears as a label next to the area for entering data. This is illustrated in the following example.

You might have the following descriptors in a passport file:

Passport

Accession number
Collector's number
Collecting institute
Date of collection
Country of collection
(other fields)

The basic entry screen supplied by the software might look something like this:

Field name	Area for data entry
ACC_NO	
COLL_NO	
COLL_INST	
COLL_DAT	
COLL_COUN	

Entering data into the above screen form could cause problems for an occasional user. But screen forms can be designed to make data entry/modification much more user-friendly. In the following sections we will look at how you can achieve this.

2.2 How to design screen forms

The design facilities for screen forms depend on the software you have chosen; some software only has basic facilities, other software is very sophisticated. The final design should depend on the preferences of the people who will be using the-system. However, there are some general guidelines you should follow.

2.2.1 Design a form for each documentation procedure

Your software will probably have generated a basic screen form displaying all fields in each data file which could be used for data entry and modification. However, you'll often only want to work with a sub-set of the total descriptors in a file. Screen forms can be designed to display only selected fields in a format that suits your data entry and modification procedures. Such screen forms are often called *masks.*

Screen forms can be designed to display only selected fields

For example, you might be updating a few descriptors in a passport file, or entering data from a characterisation trial where only a few tests were performed. You wouldn't want to hunt round the screen looking for the appropriate fields to fill in – it would be very time consuming and frustrating and the risk of making mistakes would be high. Designing a form containing only selected fields for each documentation procedure will avoid this. Do not rely on just one form for each file.

DO NOT rely on just one form for each file

2.2.2 Avoid putting too much information on each form

If you are entering data from an extensive characterisation trial and you have a single form which contains all the descriptors, the screen might appear cluttered and be difficult to read. If it's difficult to read, it's also difficult to use and you might make mistakes. You can avoid this if you distribute the descriptors over more than one form – the extra forms can be used in sequence. The screens will be easier to read and they will be easier to use.

Distribute the descriptors over more than one form

You might also want to include a list of acceptable responses next to each descriptor for users unfamiliar with your coding system, but this will take up a lot of screen space. This information could, perhaps, be put elsewhere such as in an on-line help facility or documented on paper.

2.2.3 Use descriptor names on forms, not field names

If your field names are limited to a certain number of characters, you may have had to abbreviate descriptor names to produce field names. For occasional users, it will not be obvious what the abbreviations mean or which data belongs in which field. For instance, they might think that "COLL_DAT" refers to data on "collection source". Here, data entry is clearly not user-friendly and errors can easily creep into the system.

Screen forms allow you to display more descriptive text as labels to your fields, often without a limit on the number of characters you can use.

For example, instead of: COLL_DAT
You could use the label: Date of collection

Field names can be replaced on the screen with a more explanatory label so that it is clear to the occasional user which data should go in a particular field. Therefore, it is a good idea to use full descriptor names as the field labels on your screen forms, rather than the field name.

Taking the previous example of the basic entry screen for the passport file illustrated on page 185, you could design your screen form to look like fig. 3.

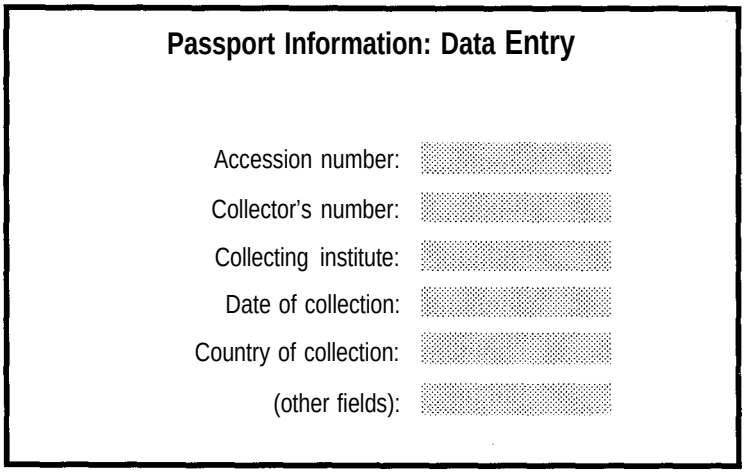


Fig. 3. Sample screen form for data entry, passport information

2.2.4 Use the same order of fields for the screen form as for the corresponding manual form

You can usually place fields **ANYWHERE** and in **ANY ORDER** on the screen

Quite often you'll be entering data into your database from manual forms. This is made easier if the fields appear on the screen in the ***same*** order as they appear on the manual form. It's very tedious and a waste of effort to rearrange the data before entering them in the database files. Remember that when you design the form, you can usually place fields ***anywhere*** and in ***any order*** on the screen; you are not restricted to the order in which the fields appear in your record definition (see Chapter 8, section 4, page 167 for full discussion of record definitions).

2.2.5 Aim for consistency in design

Put the same type of information in the same area of each form

Regular users will find the system easier to use if there is consistency in the design of different forms. Try to put the same type of information in the same area of each form. For example, error messages, information about the function of certain keys, how to get help, etc.

Give the form a title – and make it descriptive

And don't forget to give the form a title. Think about the documentation procedure for which the form is being used and make the title descriptive. If the procedure is for registering new accessions in the genebank then call the form "Register new accessions" or "Allocate genebank accession number". "Data entry form" is not good enough – data entry for which file?

2.2.6 Exploit the features of the software to minimise errors in data entry

Often the software will let you define the way in which a field on a screen form accepts and displays data. This is very helpful in trapping errors before they get into the database. Some of the more noteworthy features are discussed below.

1. Disallow invalid responses

A field can be set up so that only valid responses are accepted. This is very useful for trapping errors when you are using codes for descriptor states. To emphasise this point, let's look at an example of a characterisation file where data entered is often coded.

Characterisation

Flower colour
Flower size
Plant height
Leaflet size
Fruit firmness
Fruit aroma

For most of these descriptors, there are only a limited number of valid entries, for example 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, X. However, any of the characters on your keyboard will be normally accepted by your computer when only perhaps 12 of them are valid. You could find your system full of invalid data. This can be avoided by designing screen forms to restrict data input to valid responses.

2. Use the help message display

This can be used to give information about the field, such as what is a valid response.

3. Display error messages

This is useful for informing the user when a mistake has been made and what can be done to correct the mistake.

4. Store a default value

Fields are usually empty until data are input. However, with certain fields, you might want to store a default value until data are input. This is useful if the initial response is always the same, such as "unknown" in a characterisation trial where the test has not been performed. Before storing default values, think about whether the default values will apply in all cases.

5. Use automatically generated field values

Certain field values which are used regularly can be automatically generated by your computer, for example today's date. This is useful when inputting passport data, or any other data where you wish to record the date of input or modification. Some software have the facility to copy the same field from the previous entry at the press of a key. This is very useful when entering a series of records from, say, a collecting mission where the accessions were collected from the same site.

6. Convert to upper case

Sometimes you might only want to store data in capitals. Fields can be defined to convert all characters that are input into upper case. For instance, if you typed a donor number as "egru 380" it would store "EGRU 380". This can help subsequent information retrieval.

7. Use display only or protected fields

It is useful to be able to protect certain fields from data modification. For example, if they must be visible, but should not be modified such as accession number or batch reference. This prevents accidental modification.

All these features are very useful and are well worth exploiting to the full. They may not trap *all* the errors, but will significantly reduce the number that ret into the system.

8. Allow cursor movement around the screen

It is very useful to be able to review a screen and correct any errors before moving on to the next screen. If your software allows this feature, use it routinely in all your screen forms.

2.2.7 Use different styles sparingly

The use of different styles was discussed in detail in Chapter 6. Remember that you should use different styles to make working with the form easier. For example, too many large or bold letters, too many boxes and lines and too many colours can make forms difficult to read. Concentrate on the layout, using different styles for emphasis only and keep the form simple.

2.2.8 Sound effects

Restrict the use of sound effects. They could be used for alerting the user when the computer has completed a time-consuming task or the operator has made an error. But it can be irritating if the computer is continually making noises – irritating not only for the user, but also for anyone else who happens to be in the same room.

2.2.9 Experiment with different designs then choose the one which is the most practical

Some forms can look good, but not be practical to use. They might be difficult to read, have too much information on the screen or need better on-line or documented help facilities. It's worth producing different versions and asking users to experiment with them. With feedback from the users, you can develop a form that is easy to use. Remember, the more user-friendly a form is, the more likely it is that data stored will be reliable and accurate.

2.3 Stages of form design

1. Before starting to create your form using your database software, it is a good idea to make a preliminary sketch of the form on paper. List all the fields you want to see displayed. Consult the list of field specifications you have produced for each database file, as these will influence how you wish to lay out your form.
2. Familiarise yourself with the features that your software has for designing forms – experiment with a few designs, try out some or all

of the features listed in section 2.2.6. Find out how easy each form is to use by entering some test data.

3. Design a data entry form and, if necessary, a data modification form for each complete database file. These forms will allow you to work with all the fields in each database file at one time. If there are two separate forms, they will probably be almost identical in appearance apart from the title, but the data modification form might have a few fields which are protected or display only, such as accession number and batch reference.
4. Next, design data entry and where necessary, data modification forms for each procedure using the same principles as in point 3 above. These forms will probably contain a sub-set of the total number of fields possible for a file.

3 Developing the reports

Generating reports is one of the most valuable ways of retrieving information from a database. Reports are usually set up when information is required on a regular basis in a specific format. The retrieval format will depend on your information needs and can be totally different from the format in which data were input.

For information requests which are not regular or predictable, there are alternative ways of retrieving information from your system. For example, you could browse through each record (or a selected set of records) in turn on the screen and print them in the format they appear on the screen. If your software has the facility, you could also use the simple interactive query language (or the more complex structured query language, SQL) to carry out the search. In this case, one or more records could be selected and a report generated from them.

This section studies the development of reports, which is a several stage process. This is illustrated in fig. 4.

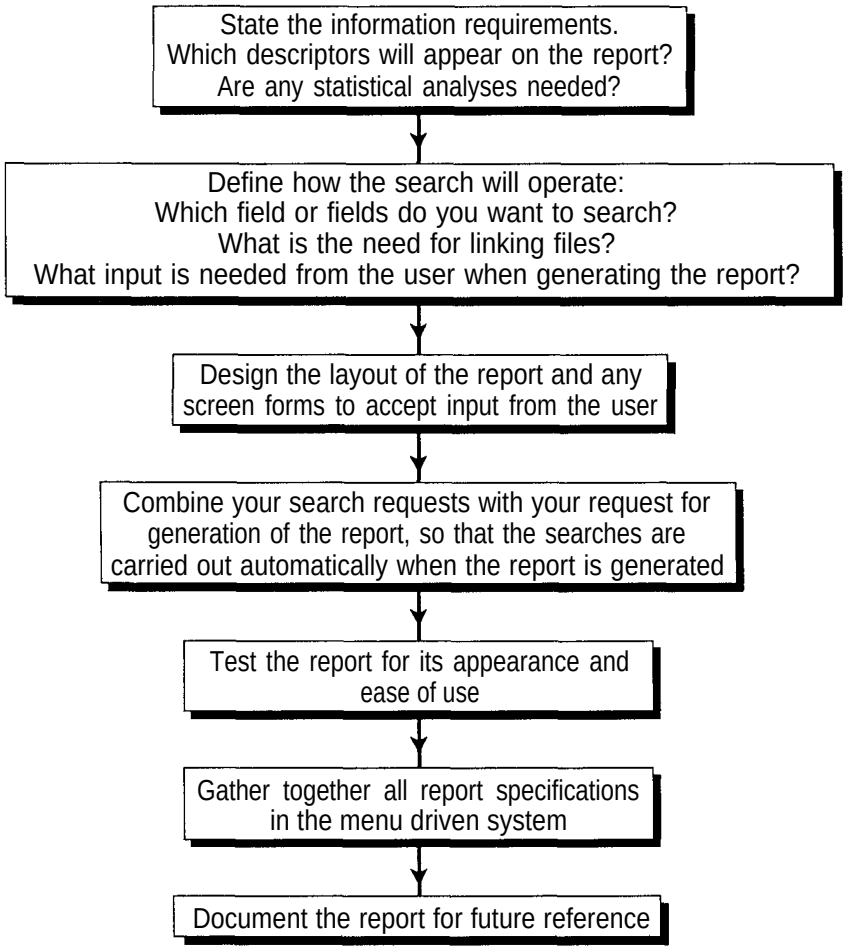


Fig. 4. Stages of report development

Some software will have more powerful reporting facilities than others. However, if you need to perform complex statistical analyses or produce graphs to get the information you require, you might need to think about exporting your data to a spreadsheet.

We will now study these stages of report development in greater detail.

3.1 Information requirements

From your genebank analysis you will have identified your information requirements.

For example, you might have talked to the officer in charge of seed testing about the information requirements for the seed testing unit and the type of reports that would be useful for the units work. You might have drawn up a list of reports rather like this:

REPORTS REQUIRED FOR SEED TESTING UNIT:

Viability of a particular accession
Viability of a particular crop
Accessions with a particular range of viabilities
Accessions that have not been tested
Accessions that require re-testing in the next 3 months
Accessions that require re-testing in the next 6 months
Accessions that are overdue for re-testing
(other reports)

For each of these reports you will have already determined which descriptors need to be listed in the report. You'll have some idea of which order the records ought to appear (such as accession number order, in alphabetical order of crop, in a date order or even a combination of any of these). You will also have identified any calculations that need to be performed on your data. Fig. 5 illustrates the idea of drawing data from different data files into a single report.

In fact, the information requirements for the different genebank activities will have already influenced some of your decisions concerning the descriptors that need to be recorded and where these descriptors are located in the data files in your system.

For instance, seed testing is performed periodically, but how often – every six months, every two years? Do some crops need testing more frequently than others? From these sorts of questions you will have identified the need to record "date of next test" which you'll use in the reports.

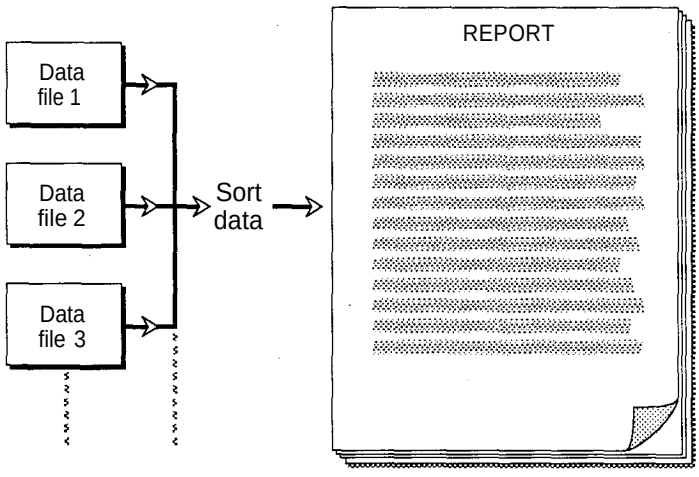


Fig. 5. Data can be drawn from a number of separate data files and reproduced in a single report

3.2 Defining the search

The next step is to develop ways of retrieving the information from the different data files in the desired order. There are three considerations which were mentioned at the beginning of this section:

1. Which fields you wish to search
2. Whether files need to be linked
3. What input is needed from the user

Let's look at these considerations with a few examples.

EXAMPLE 1: Report, in order of accession number, the seed viabilities of all maize accessions and when they need re-testing

The descriptors that you would display in this report are: accession number, batch reference, crop name, viability and date of next viability test. These descriptors could be located in two separate files: the general registration file and the seed viability testing file:

Registration	Seed viability
Accession number ◆	**Accession number** ◆
Scientific name	**Batch reference**
Crop name	**Viability**
Pedigree/Cultivar	Date of viability test
Depositor number	**Date of next viability test**
Acquisition number	(other fields)
Type of material	
(other fields)	

Key
◆ link field
Desired descriptors are **bold**

Before a report can be generated, the files must be linked. Each record in the seed viability file must be linked to a record in the registration file where they have the same accession number, so the accession number would be the identifying field. Once the link has been established, it's simply a matter of:

1. Selecting the linked records which have a crop name of "maize"
2. Sorting the linked records into accession number order
3. Printing the relevant fields in your chosen format

Your report might look something like the report in fig. 6.

NEXT VIABILITY TEST DATE: MAIZE			
Accession number	Batch reference	Viability	Date of next viability test
EGB 1007	13/07/91	80%	22/07/92
EGB 3231	15/04/90	75%	13/09/92

Fig. 6. Sample report, ordered by accession number, for seed viabilities of maize accessions indicating when they need re-testing

Note: Even though the field "crop name" is used in generating this report to specify that we want to search for records with crop name "maize", there is no need to display this field on your final report. The report is only concerned with maize, so the same information would be repeated for each record if the crop name field was included. Instead, you could make the reference to maize in the title.

EXAMPLE 2: Report, in order of crop name and accession number, all accessions which have seed viabilities in a particular range

This example is similar to the one above except that the output needs to be sorted by crop name and accession number and the user has to input a range of viabilities, e.g. 70-80s.

How does the user supply the range of viabilities? You could design a screen form which would be displayed before the report is generated, asking the user to specify the range of viabilities. In this case, the user would-specify the seed viability range 70-80%.

When the user has entered this range, the procedure is similar to example 1. The files are linked together using the accession number. Then, records are selected which have the specified viabilities. The linked records are sorted into crop name order; within each crop name group they are also sorted into accession number order. Finally, the relevant fields for each record are printed in the specified format.

On this report, you may wish to display other fields instead of batch reference and date of next viability test. The final report might look something like the report in fig. 7.

ACCESSIONS WITH SEED VIABILITIES IN RANGE 70-80%				
Crop name	Accession number	Batch reference	Viability	Date of next viability test
Barley	EGB 4962	17/07/91	79%	28/07/92
	EGB 4976	21/05/91	78%	05/06/92
Maize	EGB 1007	15/04/90	80%	22/07/92
	EGB 3231	15/04/91	75%	13/09/92
Wheat	EGB 1999	17/10/91	72%	14/08/92

Fig. 7. Sample report, ordered by crop name and accession number, for accessions which have seed viabilities in a particular range

EXAMPLE 3: Report, in order of crop name and accession number, all accessions which have not been tested for seed viability

This time we need to search for "absent" records in the seed viability file, i.e. records which have not been tested for seed viability. The link between the two files is again the accession number. You would instruct your computer to select records which do not have a corresponding record in the inventory file, i.e. they have not been tested for seed viability. The selected records are then sorted into crop name order and accession number order. Finally, the relevant fields for each record are printed in the format you have designed.

The final report might look something like the report in fig. 8.

ACCESSIONS NOT TESTED FOR SEED VIABILITY		
Crop name	Accession number	Batch reference
Barley	EGB 4812	14/09/91
	EGB 5060	27/07/91
Maize	EGB 1234	18/05/91
	EGB 2231	17/12/91
Wheat	EGB 2001	09/12/91

Fig. 8. Sample report, ordered by crop name and accession number, for accessions which have not been tested for seed viability

EXAMPLE 4: **Report, in order of crop name and accession number, all accessions which have low seed viabilities and list their seed weight in the cold store**

To access data in the total seed weight field, you might have to link three files together; the registration, seed viability and inventory file:

Registration	Seed viability	Inventory
Accession number ◆	**Accession number** ◆	**Accession number** ◆
Scientific name	**Batch reference** ◆	**Batch reference** ◆
Crop name	**Viability**	Location in store
Pedigree/Cultivar	Date of viability test	**Total weight of seed**
Depositor number	Date of next viability test	1000 seed weight
Acquisition number	(other fields)	Minimum weight of seed allowed
Type of material		(other fields)
(other fields)		

As you will probably have more than one batch for each accession, the seed viability file and the inventory file will have to be linked through the batch reference field as well as the accession number field. However, both files can link to the registration file using the accession number alone.

As in example 2, records are selected which have low seed viabilities by inputting a specified range. Alternatively they could be selected by inputting a single value below which viabilities are considered to be 'low'. The selected records are sorted into crop name order; within each crop name group they are also sorted into accession number order. Finally, the relevant fields for each record are printed in a particular format.

Your final report might look something like the report in fig. 9.

ACCESSIONS WITH LOW SEED VIABILITIES: TOTAL WEIGHT OF SEED				
Crop name	Accession number	Batch reference	Viability	Total weight of seed (kg)
Barley	EGB 4962	17/07/91	79%	0.243
	EGB 4976	21/05/91	78%	0.567
Maize	EGB 1007	15/04/90	80%	1.440
	EGB 3231	15/04/91	75%	2.678
Wheat	EGB 1999	17/10/91	72%	1.999

Fig. 9. Sample report, ordered by crop name and accession number, for accessions which have low seed viabilities showing total seed weight

In summary, for each report you will need to specify:

- How files are linked together
- The criteria for the search (e.g. a particular crop, a range of viabilities)
- How the search criteria are specified (embedded in the report routine itself or specified by the user on a form)
- How records should be displayed (sorted in a particular order, which fields will appear on the report)

You should develop similar search specifications for all the reports you want to generate.

3.3 Report styles

Designing reports should be approached in much the same way as screen forms. They should be user-friendly and constructed so that they are easy to read and to understand. It's therefore worthwhile spending time designing your reports. After all, you will be relying on them to make important decisions such as decisions about regeneration priorities. How precisely you are able to design reports will depend on the software you are using.

You could design a report which lists one accession on each page. In such a report, each descriptor will appear in exactly the same place on every page. It's similar to the method used in screen forms where each descriptor has its own specific location on the screen. You would use this style of report if, for example, you were giving detailed information on the results of a characterisation trial for a particular accession (see example in fig. 10), or using pre-printed stationery (such as invoices), or producing address labels.

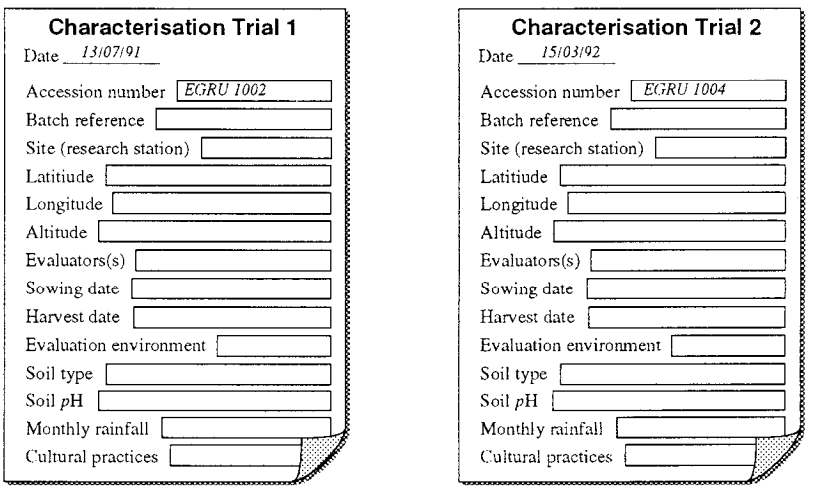


Fig. 10. Detailed reports for two different characterisation trials. Descriptors are in the same place on each page

More commonly, you'll want to produce reports which list selected data for several accessions on one page in a column format, similar to the forms illustrated in examples 1 to 4 earlier in this chapter (see figs. 6-9). Each descriptor will occupy a particular column on the page. The report might extend over several pages. Each page should therefore have the same basic design.

3.4 Design of the report

3.4.1 Give each page of the report a title, a page number and a date

You must be able to tell at a glance what any report is about. Therefore, give it a title!

Reports often extend over several pages. If the pages get separated, it may not be obvious which report they belong to or where they belong in the report. So, put the title and a page number on each page. Also, get into the habit of putting the date on the report. You don't want to mix up the old reports with newer ones, as you could find yourself working with out-of-date information.

3.4.2 Arrange the columns in a useful order

Think carefully about how the columns should be arranged on the form. If you are comparing two columns it's sensible to put them side by side.

3.4.3 Give each column a title

It must be clear to the person reading the report what information is contained in each column. Descriptors are the obvious choice for column headings. But problems occur when you have a lot of columns and the descriptors have long names such as "mean days to flowering" and "resistance to *Helminthosporium maydis*". You will quickly run out of room on your page.

The easiest solution is to wrap column headings onto more than one line, so you should first explore whether your software allows you to do this. Otherwise, if the software you are using has a "report summary" feature, you can give each column a number or abbreviation and then in the report summary list the column numbers or abbreviations next to the relevant descriptor name (see fig. 11). If your software has neither of these features you could consider:

1. Splitting the report into two or more separate reports containing sub-sets of the descriptors. If you do this, think carefully about what information you need to see on the same page.

2. Producing the report with column numbers or abbreviations and listing these on an separate information sheet. If you use this option, you should ensure that you supply the information sheet with each report otherwise there might be some confusion.
3. Think again about how the report is going to be used. Do you *really* need all the specified descriptors on the same page?

ACCESSIONS WITH SEED VIABILITIES IN RANGE 70-80%				
1	2	3	4	5
Barley	EGB 4962	17/07/91	79%	28/07/9
	EGB 4976	21/05/91	78%	05/06/9
Maize	EGB 1007	15/04/90	80%	22/07/9
	EGB 3231	15/04/91	75%	13/09/9
Wheat	EGB 1999	17/10/91	72%	14/08/9

1. Crop name
2. Accession number
3. Batch reference
4. Viability
5. Date of next viability test

Fig. 11. Report with numbered columns. Column numbers are listed next to corresponding descriptors in the report summary

3.4.4 Aim for consistency between reports

People will find it easier to use reports if there is consistency in the design of different reports. For instance if you are producing different reports on seed inventory it will help if the columns are arranged in the same order.

3.4.5 Exploit the available features of your software

Study all the features of the software you have chosen and exploit them to the full in your reports.

For example, certain software is able to group records with the same value for a particular descriptor (e.g. the same crop name) and then perform calculations or statistical analyses on other fields in the record. So, in a report on accessions which need seed testing it could calculate how many accessions of each crop need testing, or even how many accessions need testing in a particular month. This kind of information can be very useful and might save you a lot of time in the future.

3.5 Printing the report

"The report looks alright on the screen but when I print it out it looks terrible!"

If you have used other software for word processing, desktop publishing, graphic design or spreadsheets, you may have encountered problems when trying to print out your work. Many of these problems can be solved after a careful reading of your printer manual and the section in your software manual concerned with printing the reports. However, there are a number of problems which are commonly experienced which are discussed in the following sections.

3.5.1 The size of characters and the printer font

Your printer might be able to print characters in different sizes. Obviously, when the character size is larger, fewer characters can fit on a line. It might also have the ability to print in a number of different fonts. Compare the following three different fonts:

Font 1: Plant Genetic Resources

Font 2: Plant Genetic Resources

Font 3: Plant Genetic Resources

Notice the difference between the same character in different fonts. Notice also that some fonts take up more space than others so the line length for "Plant Genetic Resources" is different in each case.

Fonts can be split into two. categories according to how they are spaced: monospaced and proportionally spaced. Each character in a monospaced font occupies the same space e.g. "i" takes up the same amount of space as "o". With proportionally spaced fonts, the space occupied varies from character to character. The difference between the two classes of font can be demonstrated by the following example:

Monospaced font:	ooooo iiiii	ooooo iiiii	ooooo iiiii	ooooo iiiii
Proportionally spaced font:	ooooo iiiii	ooooo iiiii	ooooo iiiii	ooooo iiiii

Check the fonts available to you in your software and on your printer and ensure that the settings do not conflict.

If the report columns are aligned on the screen but not on the printed page, you should:

1. Check the font that is being used. Try changing this to a monospaced font such as “Courier”
2. Check the character size. Is it too large? If so, reduce it. Consult your manuals to find out how to do this.

3.5.2 Limitations of some printers for report printing

Dot-matrix printers can only display a limited number of characters per line. Many dot-matrix printers have 2 paper size settings; standard, which allows 80 monospaced characters per line, and wide which allows 132 monospaced characters per line. If the total width of fields in a report exceeds this character limit (including the spaces), extra text will wrap onto the next line. Therefore, if you are using a dot matrix or daisy wheel printer, make sure that the total width of your report does not exceed this limit.

If you are using wide paper make sure that the printer has been set up to accept the wide paper: the printer is not intelligent – it has to be told to expect wide paper, either by the software or by changing the manual settings on the printer.

3.5.3 The page length

The page length for your report must be set in your software and, for some printers, it also needs to be set on the printer itself.

It is often possible to alter the settings on your printer, so that it can accept a certain length of paper (e.g. 8.5", 11", 11.7" (A4), 12"). If the printer is set to expect 12" paper and you use 8.5" paper, the results can be unpredictable. Make sure that the settings on the printer agree with the paper size you are using. Consult your printer manual to-see how you can configure the printer correctly. Bear in mind that sometimes the software can overrule the settings on the printer.

If you have to specify the number of lines per page in your software when you design a report, make sure this corresponds to the correct number of lines for the size of paper you are using. If you make this number too high or too low, your pages might have large areas of blank space.

3.6 Stages of report design

The stages of report design are summarised in the chart in fig. 12.

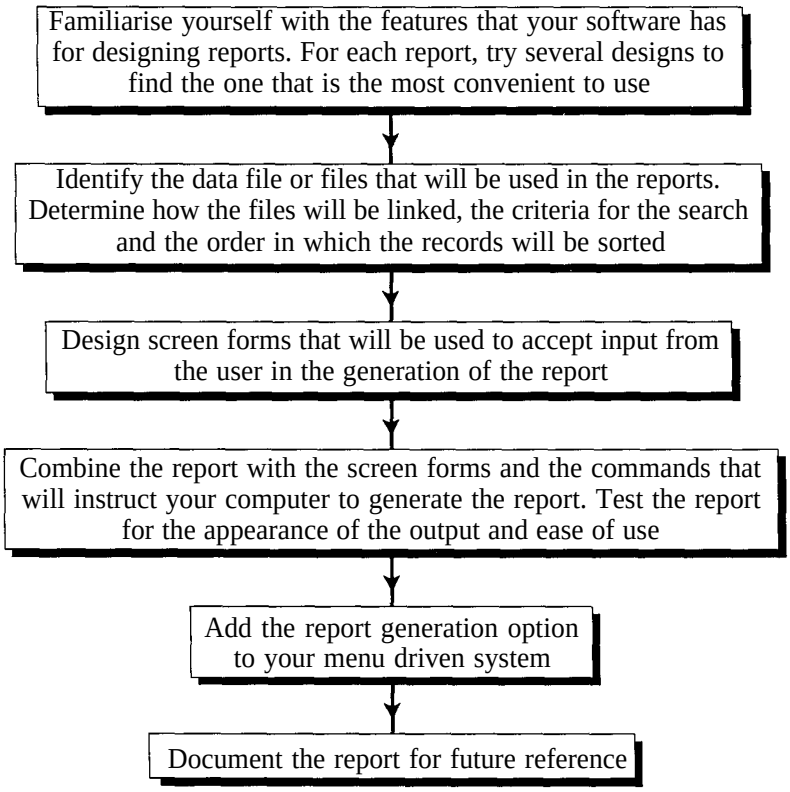


Fig. 12. Stages of report design

4 Implementing the data management features of your software

Once you have designed the various screen forms and report formats, you can begin to work on the routines for managing the data in your documentation system. It's not possible here to go into great detail

about how you can do this, as a lot depends on the software you are using but we can look at some general considerations.

The stages of developing your data management routines are illustrated in fig. 13.

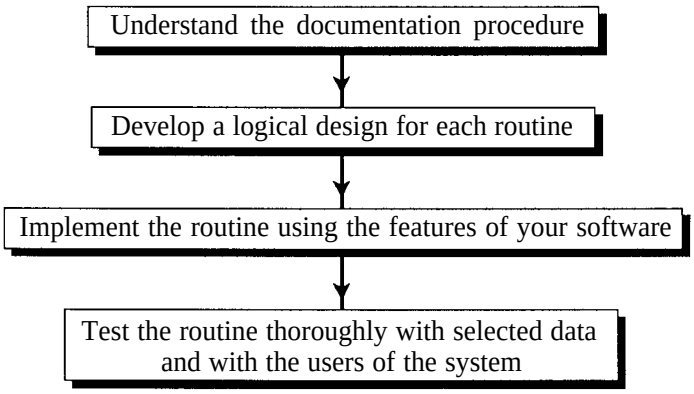


Fig. 13. Stages of developing data management routines

In the following sections we will look at some considerations which might affect the way you manage your data.

4.1 Data entry (new records)

Usually when entering data, you will only be working with one file. However sometimes you will want the data you enter in one file, to automatically trigger data entry, or automatically update data in another file.

For example, when seeds are removed from the cold store, the corresponding documentation procedure would involve:

1. Entering a new record in the movements file
2. Updating the appropriate record in the inventory file.

Using your software, you can set up a system to automatically update the appropriate record in the inventory file when you enter a new record in the movements file. But before this can happen, the files must be linked.

There are some other considerations when developing routines for data entry.

4.1.1 Trap duplicate records

You can use your software to disallow duplicate records. The following examples should illustrate how this can be achieved.

- Registration and passport files

There should only be one record for each accession since the registration and passport data will *always* be the same for a particular accession. Therefore, when these data are entered, a check should always be made to see whether a record already exists for the accession. If a record *does* exist, either data entry should be disallowed, or the user should only be allowed to *modify* certain fields of the existing record.

- Characterisation files

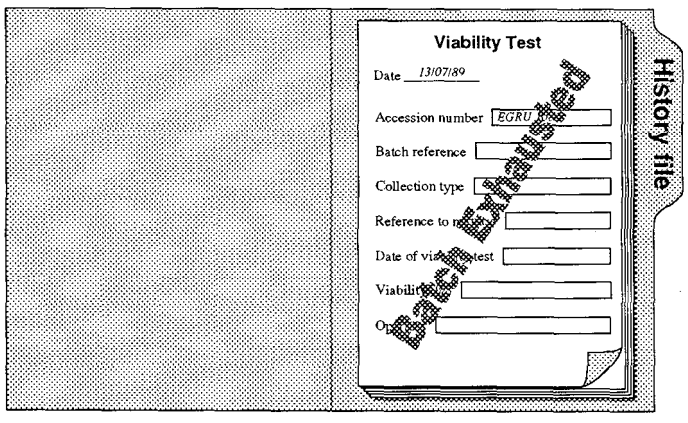
Duplicate records can be trapped by cross-referencing the identifying fields of accession number, batch reference and date of test. These uniquely identify a record. If a record exists, either data entry should be disallowed or the user should only be allowed to modify certain fields of the existing record.

- Inventory file

The identifying fields of accession number and batch reference can be used to trap identical records. These uniquely identify a record unless identical seed batches are stored in different locations in the cold store. Again, if a record. exists, either data entry should be disallowed or the user should only be allowed to modify certain fields of the existing record.

- Seed viability testing file

You can trap records which have the same accession number, batch reference and date of test as these uniquely identify a record. However, over a period of a few years this file might become very large and it will contain test data for seed batches which no longer exist in the cold store. How can you ensure that this file only contains data on seed batches which exist? You could consider implementing an additional routine for the maintenance of the inventory file that will allow the user to do two things:



1. Delete a record from the inventory file. i.e. a particular batch of seed which has been exhausted
2. Remove all seed viability testing data relating to this *batch* to a second seed viability "history" file. This history file might be useful some time in the future as an information source for, say, the long term viabilities of seeds.

Before deleting a record, you must think carefully about whether you are likely to want to use the data again. If you think it might be useful for future reference, it should be moved to the history file. Otherwise, it can be deleted.

4.1.2 Use the software to allocate accession numbers

Two accessions should **NEVER** have the same accession number

It's vitally important that two accessions should never have the same accession number. After all, the accession number forms the basis of the entire documentation system for accession-specific data!

If the allocation of accession numbers is part of the documentation procedure, you should use the software to assign them automatically. You would then never have to enter a new accession number, it will be done for you.

If your software does not have an automatic option to do this, you could consider keeping a separate data file which maintains one record – the "most recent" accession number. Let's call this file the "count" file. When the screen form for data entry is used, the "most recent" accession number is increased by 1 and displayed on the new form. After successful data entry, the incremented accession number would be stored in the "count" file.

Can you think of any occasion when you might want to enter the accession number manually in the registration file? You would want to do this when you are changing from a-manual to a computerized system and you need to enter registration data for existing accessions. However, you must make sure that duplicate records are still trapped.

So, you will have two routines for data entry (new records):

1. Sample registration – allocate new genebank number
2. Enter registration data on existing accessions

4.2 Data modification

You should use the features of your software to set up systems for data modification for each data file in your system. These can be used to correct errors in specific records, update records with new data and delete any unwanted records. The following features could be incorporated:

1. *Browse facility:* This allows you to work through one record at a time and view/edit each record. Usually you can specify the order in which you want to see the records by sorting before browsing, e.g. in increasing order of accession number.
2. *Search facility:* This allows you to locate a particular value in a specific field. For example, you might want to locate a particular accession number.
3. *Protect certain fields:* This ensures that fields which the user should not be able to modify are "read only" e.g. the accession number.

4.3 Reports

We have already covered in some detail the design and operation of reports in an earlier section. The separate elements of your report routine, i.e. the report design, the search criteria and the user input must be integrated into a single routine option that can be chosen from your menu. Make sure you include options to:

1. Display the report on the screen
2. Write the report to a file
3. Print on paper/pre-printed stationery

4.4 Examples of logical design for data entry in a crop passport file

Illustrate the stages of each routine as a logical design

Before you begin to work with the data management facilities of your software, it is very useful to illustrate the stages of each routine as a logical design. Consider a procedure for the entry of passport data for a particular crop. The aim of the procedure is to accept and store passport data for one or more accessions of a particular crop. A possible logical design is illustrated in fig. 14.

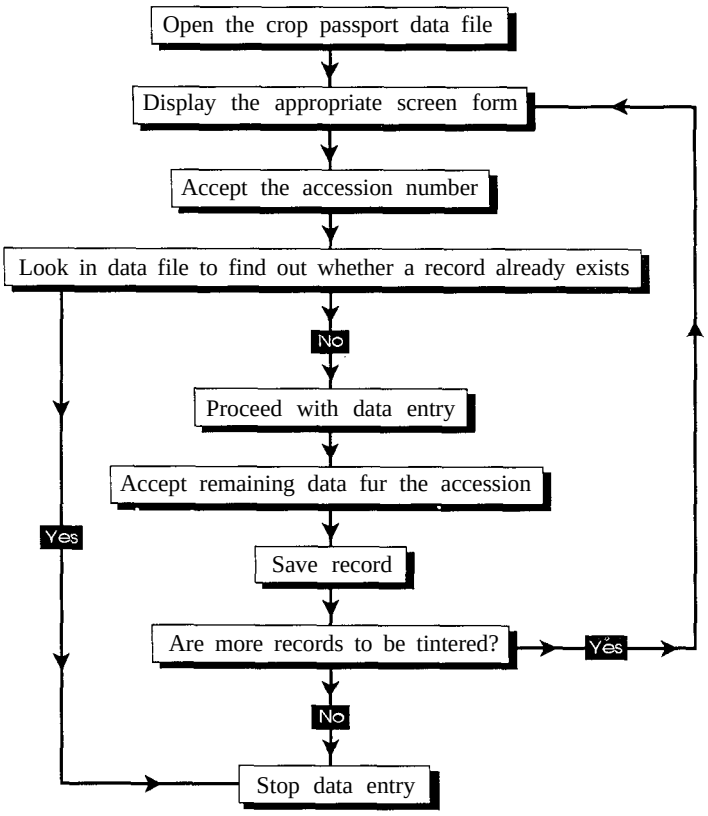


Fig. 14. Example of logical design for data entry into a crop passport data file

However, this is not the only logical design you could develop. Consider the alternative logical design in fig. 15.

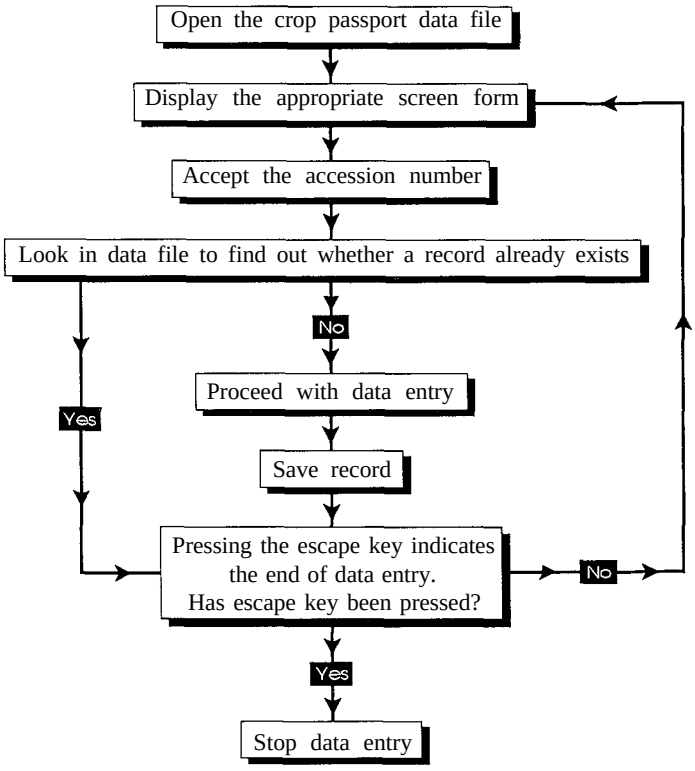


Fig. 15. Example of alternative logical design for data entry into a crop passport data file

Illustrating the steps of the procedure in this way will help you enormously when you come to implement each routine using your software.

5 Developing the menus

To help users in their selection of procedures, all your routines for data entry, data modification and report generation should be brought together into a series of user-friendly menus. Well designed menu driven systems are easy to use and they're a familiar feature of many commercial software. Let's look in more detail at how you can design them.

5.1 Menu organisation

Keep the number of items on a menu to a manageable number

If you have a lot of different routines, it is a good idea to organise them into smaller groups for convenience, each group occupying a sub-menu. You'll do this to keep the number of items on a menu to a manageable number – the more items there are on a menu, the more difficult it is to use. But how do you organise these groups? There are two approaches to consider.

5.1.1 By data operation

You can design the menus according to the data operations, for example, enter, modify, search, report. The "enter" sub-menu would contain the data entry procedures, the "modify" sub-menu would contain all the data modification procedures and so on.

For example, suppose you want to enter characterisation data for Wild *Arachis.* You could arrange the first menu ("main menu") to offer a choice of data operations, the second menu to offer a choice of subject areas and the third menu to offer a choice of crops. The sequence might look like the series of menus in fig. 16.

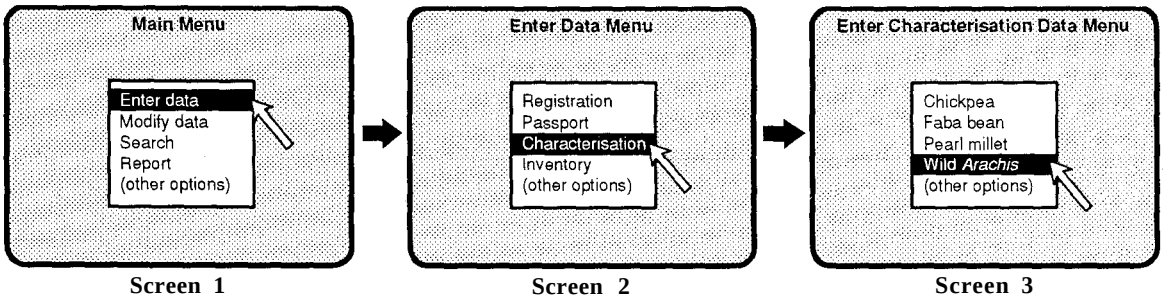


Fig. 16. An example series of menus for entering characterisation data for wild *Arachis*

You select "Enter data" from the main menu. This presents you with another menu, the "Enter data menu". From this you select "Characterisation" which presents you with the "Enter Characterisation data" menu. From the selection of crops you choose "Wild *Arachis*".

This organisation is a logical one but it assumes that the user knows what the data operation is for a particular genebank procedure. For instance, when seed is removed from the seed store, what is the data operation – "enter", or "modify"?

5.1.2 By subject area

The other approach is to design the menus according to their subject area, for example registration, passport, characterisation, seed testing, inventory. This way the occasional user knows where to look for a particular documentation procedure.

Using the example of data entry for wild *Arachis*, the sequence of menus might look something like the series of menus in fig. 17.

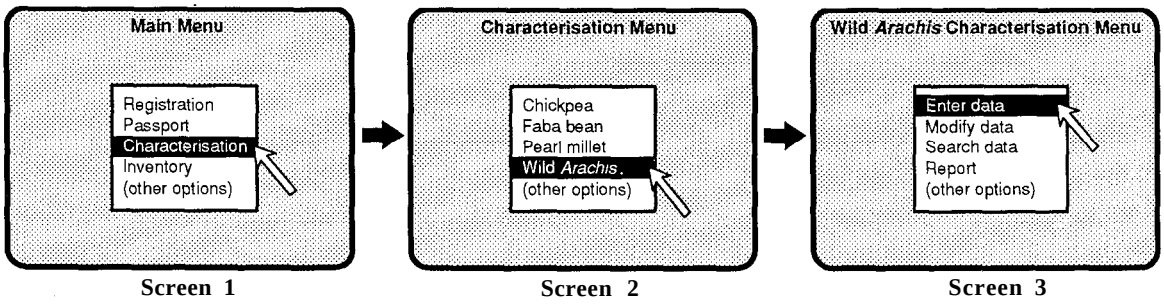


Fig. 17. An alternative example series of menus for entering characterisation data for wild *Arachis*

In both menu systems, you have to look at three menus to find the relevant procedure. Which organisation do you prefer? Which organisation do you think the users of the system would prefer? Can you think of any other organisations?

5.2 Design of the menu driven system

You may be working with one of three different styles of menu driven systems: full screen menus, bar menus and pull down menus.

5.2.1 Full screen menu

Menu options are displayed vertically, and usually centered on the screen as illustrated in fig. 18. The user selects the desired option by using the keyboard keys or mouse to select the option.

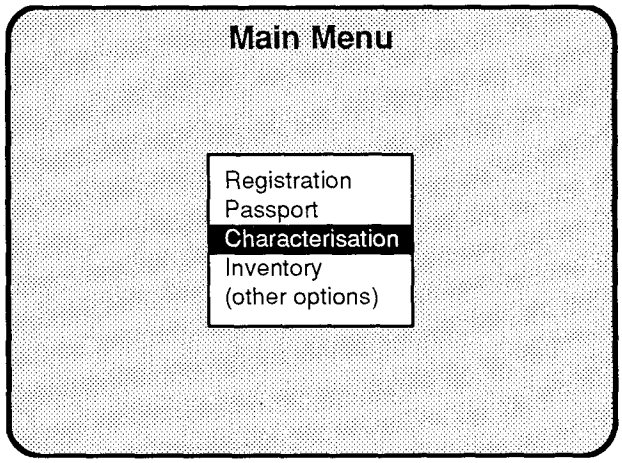


Fig. 18. Full screen menu

5.2.2 Bar menu

Menu options are displayed in a band at the top of the screen as illustrated in fig. 19. Again, the user selects an option using the keyboard keys or the mouse.

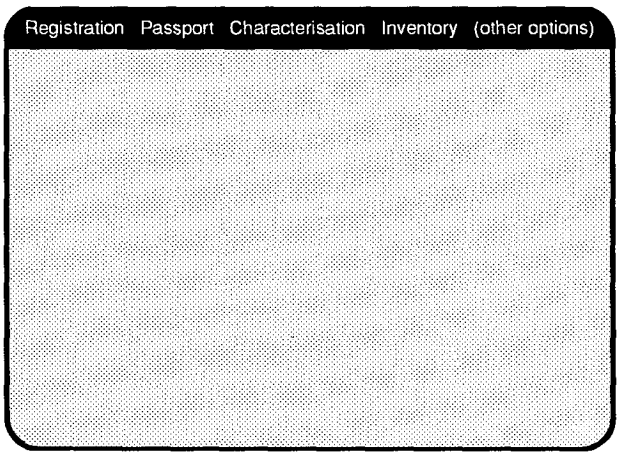


Fig. 19. Bar menu

5.2.3 Pull down menu

As with the bar menu, the user selects one of the options from this menu with either the mouse or a specific keyboard key. The sub-menu is then displayed as a column below the main menu option as in fig. 20.

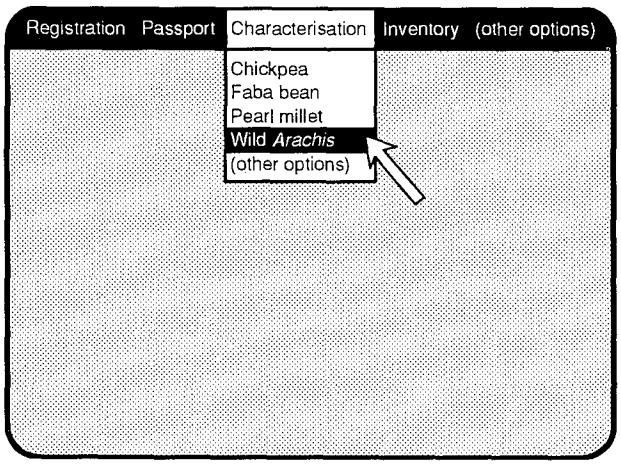


Fig. 20. Pull down menu

Bar menus are the simplest in design but you have little freedom in the layout of the screen. Pull down menus are usually faster to use, particularly when using a mouse. Full screen menus offer the most freedom in design but are often slower to use than pull down menus.

We've already looked in some detail at the design considerations for screen forms. Similar considerations apply to the design of menus. When designing any type of menu you should:

1. Organise the menus according to data operation or subject area
2. Aim for consistency in design. Experiment with different basic designs and choose one which will be used for all menu screens
3. Use different graphic styles sparingly
4. Avoid putting too many options or too much information on each menu

6 System documentation

It is important that your documentation system itself is well documented. Good documentation assists the every day operation of the system and is essential for any future modifications of the system.

Someone else might be responsible for the maintenance of the system in future and they will need to be able to understand the logical structure of the system. If you need to modify the system at a later date yourself, you will need a detailed reference source to consult: you don't want to find yourself saying "Now why did I do it like that ...?", a common question asked by people designing computer systems who have not documented their work adequately.

Each procedure should be documented as it is designed and implemented

One of the reasons that systems often have very little documentation is because documentation is often left until after the complete system has been implemented. As documentation systems are dynamic and change over a period of time, it is difficult to identify the point when they are "completed". This dynamic nature of the system often leads to a lack of documentation. Don't fall into this trap. Each procedure should be documented as it is designed and implemented. Documentation is much easier if it is done as the system is being designed and built, while the concepts are still fresh in your mind.

The approach to documentation is essentially the same for both manual and computerized systems.

6.1 Construct flow charts to illustrate each procedure

When documenting each documentation procedure, it is useful to construct a flow chart. List what occurs at each stage of the procedure. Indicate what files, menus and manual or screen forms are being used at each step and any relations between data files. Accompany each flow chart with a written description of the procedure.

6.1.1 Genebank procedures

1. Construct a flow chart for each procedure. An example was given on page 64, Chapter 4, of a flow chart illustrating the seed cleaning procedure for seeds with a high moisture content. This is illustrated in fig. 21.

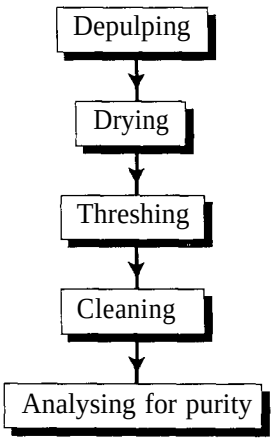


Fig. 21. Seed cleaning procedure for seeds with high moisture content

2. Construct a flow chart showing the relationship between different procedures. Which procedures are dependent on others? Which are independent?

An example showing the relationships between different procedures performed by genebanks handling seed collections (discussed in Chapter 4) is illustrated in fig. 22.

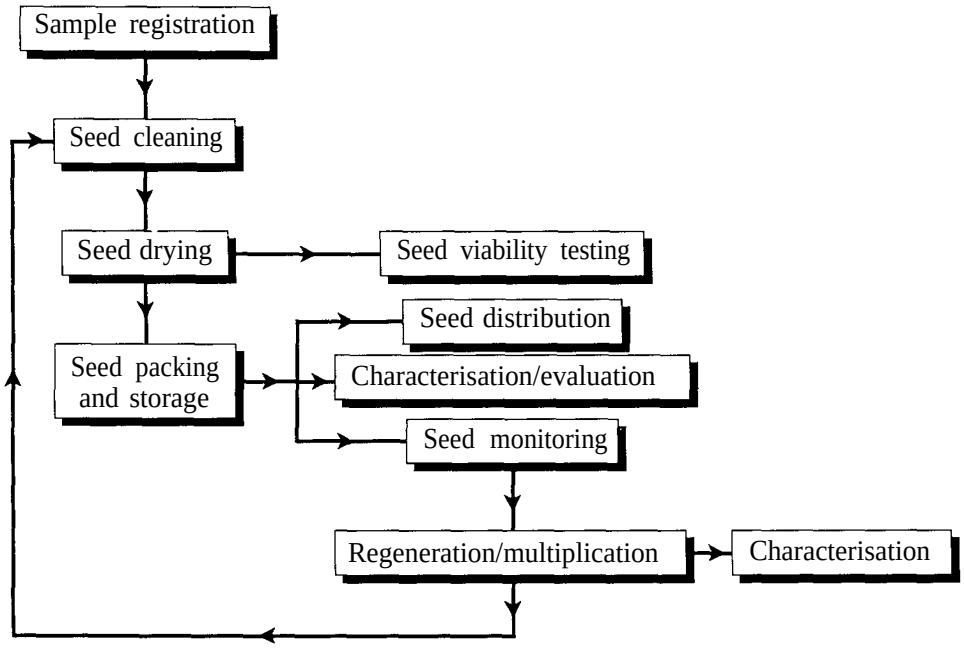


Fig. 22. Procedures commonly performed by genebanks handling seed collections

6.1.2 Data management procedures

1. Construct flow charts for information retrieval. A flow chart showing the different stages of report design was illustrated earlier in this chapter (fig. 12):
2. Construct a chart showing the organisation of the menu system. Include all the different menus and illustrate how they are related. An example is not given here as the organisation of your menu system will be specific to your requirements.
3. Construct flow charts to illustrate the processes of backing up data files and data transfer. Back up methods will be specific to your system and so are not illustrated here. Data transfer is discussed in Chapter 10.

6.2 Supply a detailed specification for each computer file

As mentioned in Chapter 8, most database management software can automatically generate a detailed specification for each computer file, referred to as a data dictionary. A data dictionary gives a listing of all the fields in a particular database file and includes details of specifications for each field. Consult Chapter 8 for further details.

6.3 Provide an explanation of all the data management features used

These will be the routines developed using the data management features of your software and should be supplied on a floppy diskette as well as in printed form. Make sure it's clear to which procedure or menu item the commands refer. Include a logical design for each option to help in comprehension.

6.4 Provide copies of all forms and sample reports

Copies of manual forms, screen forms and sample reports should be supplied with the system documentation. You should not have to consult the documentation system each time you need to see a form or report.

6.5 Make the documentation useful – include aims and objectives

It's useful if the documentation contains a list of aims and objectives for the genebank and a short discussion of how the documentation system fulfils these needs. It could also list any limitations of the current system and areas for future development.

6.6 Write a user guide

If a number of different people are using the system, it will be useful to write a system user guide. The user guide need not be as detailed as the main system documentation but instead be a simple, step by step guide to the use of the system. Since the user guide can be used for training people in the use of the system, it is discussed in further detail in Chapter 10.

7 Where to go from here

After establishing whether the software and hardware you have chosen are capable of supporting your logical system design (the developments of which were covered in detail in previous chapters), you can begin to build your system. This chapter has concentrated on the way that your system will actually operate with your chosen hardware and software, the physical system design. The chart in fig. 23 summarises the different stages of building your system.

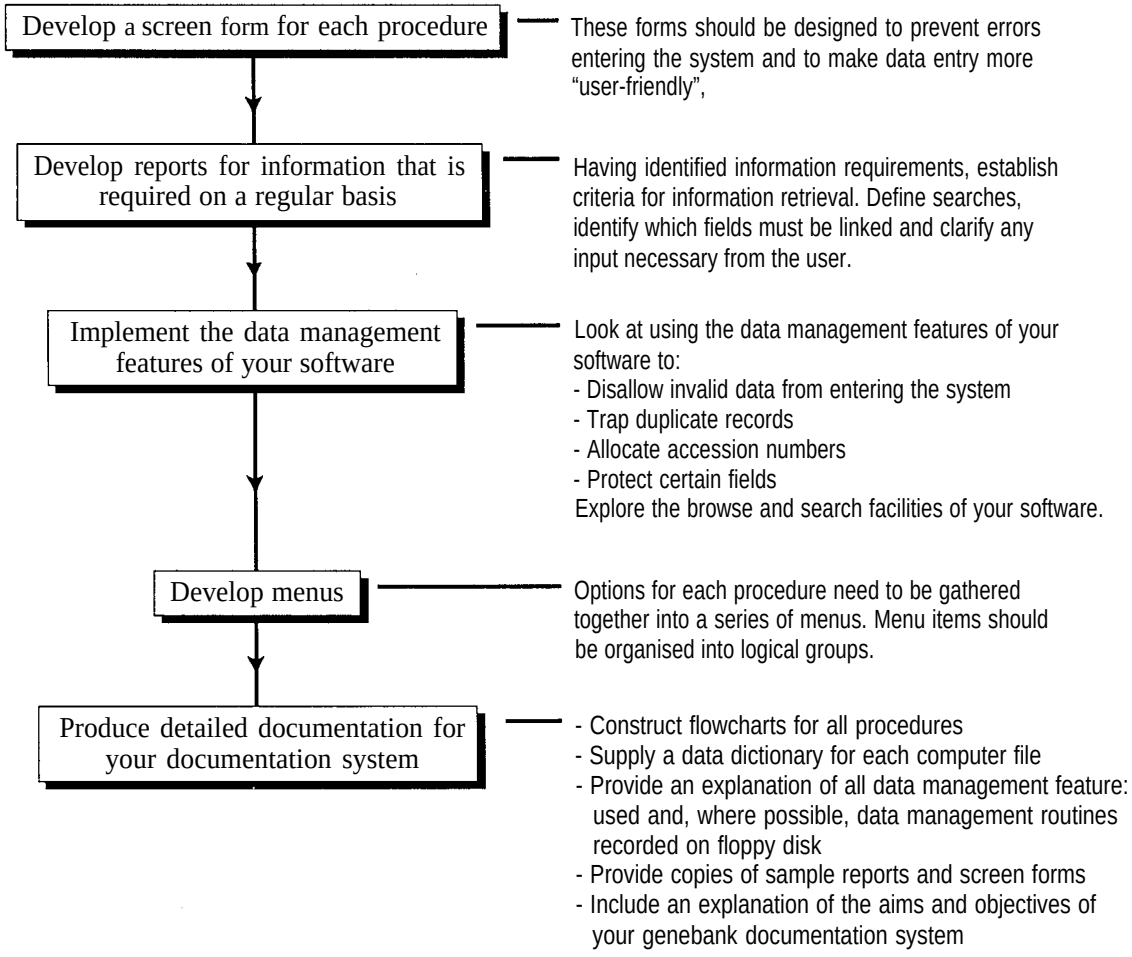


Fig. *23.* The different stages of building the system

In the next chapter we will move on to discuss the implementation, operation and maintenance of your documentation system.

8 EXERCISES

Exercises

1. Indicate whether the following statements are true or false:
   a. A logical system design is a detailed description of how the documentation system will operate using the chosen hardware and software
   b. Screen forms are used to make data entry and modification "user-friendly" and help prevent errors getting into the system
   c. Fields should appear in the same order on the screen form as they do in the computer file
   d. Features of the software should be used to trap errors after they have got into the system
   e. The format for a report can be completely different from the format of the corresponding data entry form
   f. Records should be sorted into a certain order before they are printed in the report
   g. Reports should only be printed using a monospaced font
   h. Menus make the documentation system easier to use
   i. Documentation of your documentation system should only be carried out when the system has been fully tested and implemented
   j. Flowcharts form an important part of system documentation
2. Distinguish between logical and physical design. What considerations might make you change the logical design for the system?
3. What are the advantages of using screen forms for data entry and modification? Explain how you would make a screen form user-friendly.
4. Describe the features that many software have in minimising errors in data entry.
5. Suppose you need to produce a report listing accessions with low seed viability. Describe the different stages you would follow in producing this report.
6. Describe how data management features of software can be used to build routines for data entry and data modification. Recall the logical design of a routine used for entering passport data.
7. Explain why menu driven. systems are commonly used and how the different routines in plant genetic resources documentation can be organised in a series of menus.
8. Why is it important that the documentation system itself is well documented? Describe the approaches you can use in system documentation.

Chapter **10**

Implementation and maintenance of the system

Chapter 10 will discuss the implementation of your documentation system, data exchange between different systems and applications, considerations for data security and procedures for modifying your system. When you have finished this chapter, you will be able to:

- Prepare a suitable training program for users of the system
- Select the best way for your genebank to introduce the new system
- Establish efficient routines for exchanging data:
  (i) between your documentation system and other documentation systems
  (ii) between your documentation system and other applications
- Identify the principal sources of data corruption and how these can be avoided
- Take steps to protect your system against infection by computer viruses
- Modify your documentation system to accommodate the inevitable changes in the working practices in your genebank
- Review the changing information needs of your genebank and anticipate these in your documentation system

1 Implementation of the documentation system

Implementing your new documentation, system involves drawing together the work you have done to date on all aspects of the documentation system design, to form a fully operational physical system that meets the specification developed in your analysis.

Whether your new system is manual or computerized, its implementation is a big project which requires much careful planning. If you are working in a small unit and you alone are responsible for documentation, implementation is fairly straightforward – having designed, checked and documented the system,. you can start using it immediately. The process is more complicated if you are working in a larger unit where other people will be using the documentation system, or if the documentation system will be operating on more than one site.

For your documentation system to operate successfully, it is important that someone accepts responsibility for its supervision and management. It is preferable if one person has sole responsibility for this, if it is practical at your genebank. This person (called a *system manager)* must be available to respond to comments and suggestions from users of the system on its efficiency and ease of use, and to deal with any problems that might crop up. The system manager should also be able to respond to additional information requests, whether they come from colleagues. at your genebank or from researchers at other genebanks or institutions who require information on accessions stored at your genebank.

In the following sections we will look at two areas of implementation which need to be carefully considered to ensure the success of your documentation system, namely:

1. Training users of the system
2. How you introduce the new system

1.1 Training

It is vital that anyone who will be using the documentation system is given proper training. If users are not given adequate training, they will not know how to use the system in the most efficient way, and mistakes will inevitably be made. Stumbling upon the capabilities of a system by trial and error is a very time-consuming way to learn.

Stumbling upon the capabilities of a system by trial and error is a very time-consuming way to learn

Training must be an essential part of the implementation of your documentation system, not an optional extra. It is important to think carefully about which people will be using the system, what knowledge they have already and what else they will need to know before they can carry out the documentation procedures.

1.1.1 Identify staff and skills available

Before you begin to assemble a training program, it is essential to evaluate the resources available. Staff skills and experience are a fundamental resource, which can be assessed using the chart in fig. 1.

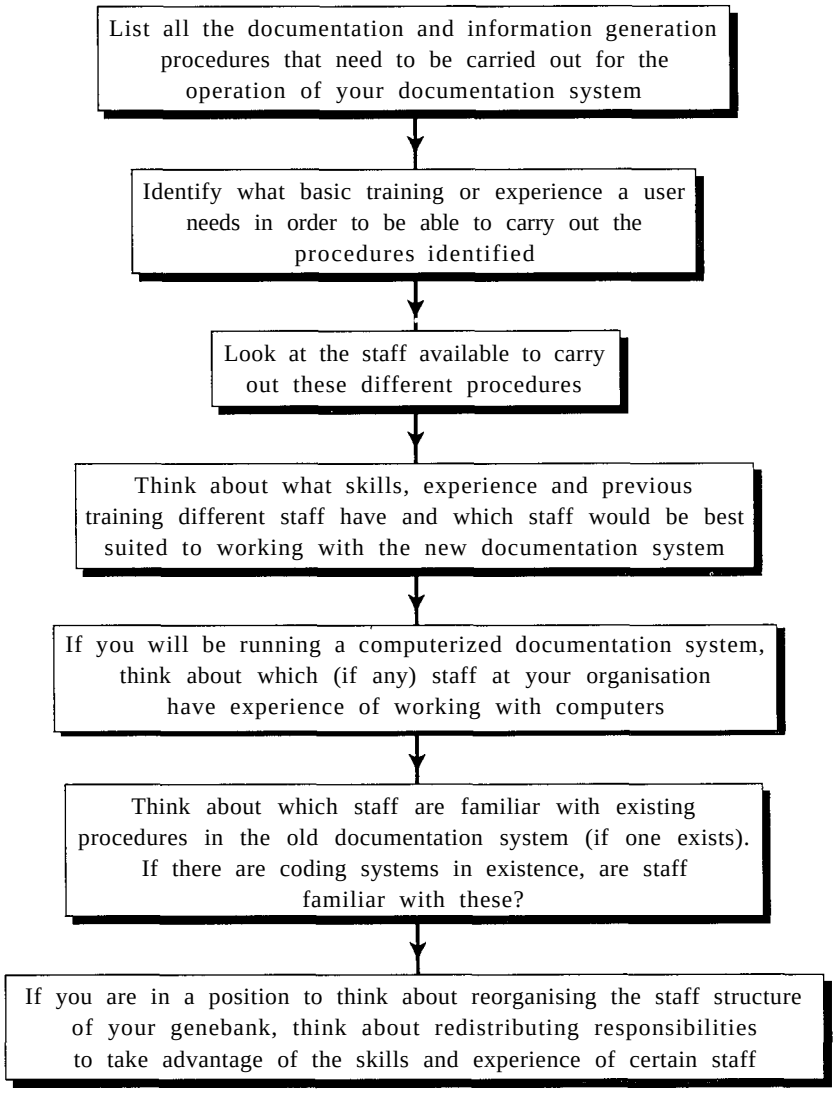


Fig. 1. Stages of evaluating available staff resources

At this stage you should also think about staff turnover. For example, if a member of staff is only going to be with your organisation for a short period of time, it may not be worth spending a lot of time training them, only to have to train another staff member after they depart.

1.1.2 Identify training objectives

Identifying training objectives involves a detailed study of each documentation procedure in turn. Here, you need to think about what the

user needs to know in order to be able to carry out the different stages of each procedure. Considerations include:

- What menu options are needed to select the documentation procedure
- How the screen forms for data entry/modification should be used
- Any coding systems that have been used
- Any scales that have been used
- Any key-strokes which are needed to carry out particular functions

An analysis should be carried out for each procedure in turn, including the options for report generation, system back-up and data exchange.

1.1.3 Assess training needs

It is vital to assess the training needs of each individual user. Look at the knowledge, skills and experience of each user in conjunction with the training objectives identified above and begin to determine what training is needed. Identify any weaknesses in staff experience which can be remedied by the relevant training.

If you're implementing a computerized system, training will need to include general instructions on how to operate the computer. The amount of training required in this area will depend on whether users have previous experience of using computers. It may be necessary to give basic instruction, covering such areas as how to switch the computer on and off, how to operate the printer and so on.

Think about whether staff will respond better to training "on the job", or whether it would be more beneficial if training is carried out away from the normal working environment. In a busy brganisation, it is often better to carry out training away from the usual distractions and interruptions of everyday work.

1.1.4 Construct the training programme

Training should be designed to stimulate interest and enthusiasm in the people who will be using the new system. Giving a system demonstration is often a good way to start. It is a valuable way of illustrating how procedures should be performed and is also an interesting way to learn. The aim should always be to give the user confidence; if they feel confident with the system, they will work more effectively and make fewer mistakes.

Think about a logical order for covering the different aspects of operation of the system. It is usually a good idea to introduce to the system as a whole, but then to cover each procedure in detail as a separate module of the training program (see fig. 2). Include a module on general

maintenance of the system, giving instructions on backing up the system and guidance on data security and how to attempt to keep the system virus free.

Training Programme Schedule

Module 1: System demonstration	All staff	Monday 2 November 10.00a.m - 12.00am
Module 2: Computer basics	Group A	Wednesday 4 November 10.00am - 12.00am 2.00pm - 4.00pm
Module 3: Procedures	Groups A & B	Friday 6 November 10.00am - 12.00am 2.00pm - 4.00pm
Module 4: Report generation	Groups A & B	Monday 9 November 10.00am - 12.00am
Module 5: System maintenance 1. Data transfer 2. Backing-up the system	Groups A & B	Wednesday 11 November l0.00am - 12.00am 2.00pm - 4.00pm
Module 6: Practical session & exercises	Groups A & B	Monday 16 November 10.00am - 12.00am 2.00pm - 4.00pm
Module 7: Evaluation, follow-up training & open discussion	All staff	Tuesday 15 December 10.00am - 12.00am 2.00pm - 4.00pm

Note: Staff members have been split into 2 training groups: A and B, based on previous computer experience. Please consult the system manager for information on grouping.

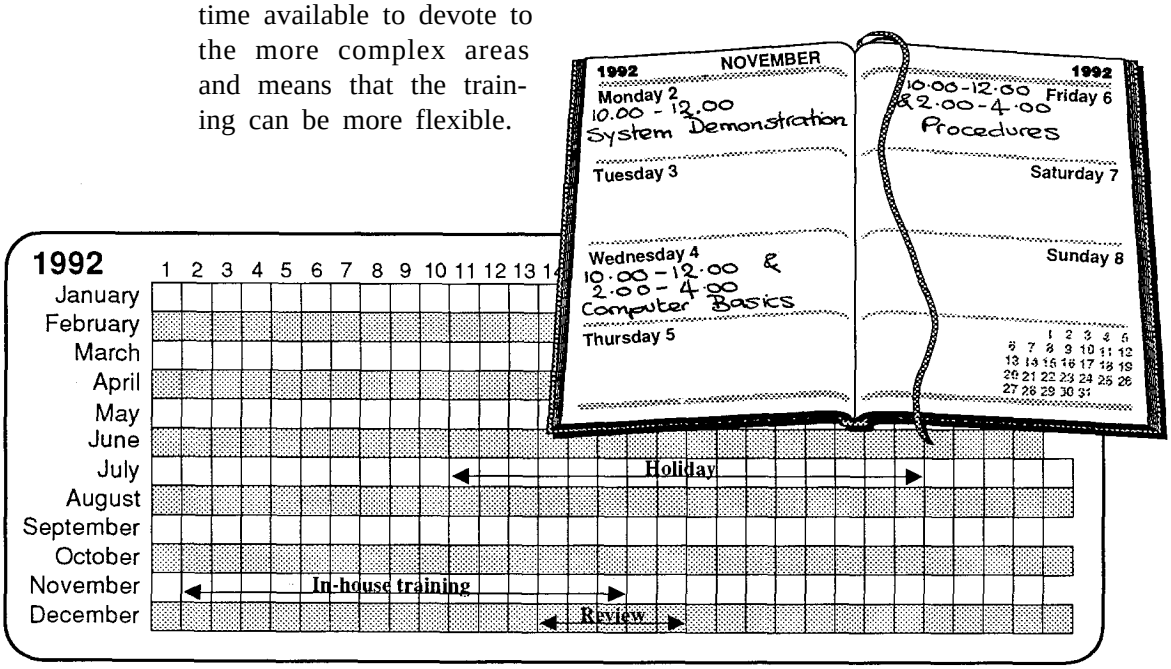
All training sessions will be held in room 216

Fig. 2. Example of a training programme schedule

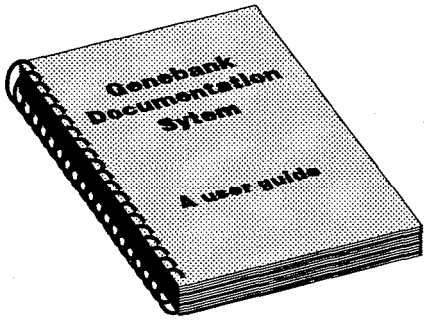
Think about whether the training should be carried out all at once, i.e. a continuous period of intensive training, or whether it would be better to break the training down into smaller sessions, spread over a longer period of time. There are advantages and disadvantages to each of these approaches. When split into smaller sessions, training does not interfere as much with the normal work load of your genebank. However, many people respond better if training is not fragmented, as it means that they can concentrate on understanding how the new system works, without the inevitable distractions of every day work.

Exercises modelled on actual documentation procedures are a useful way to give confidence in using the system. These could be carried out as part of the training program under your guidance. Make sure you set up a copy of the main database, or a database containing sample data for the user to work with at this stage. You do not want any of the accurate data in your main database to be accidentally modified or damaged in any way.

Training is more effective if performed on a personal basis. People respond well to individual tuition or teaching in small groups as they feel more involved with the system they are using. It also gives them the opportunity to raise questions on areas they do not fully understand without embarrassment or the worry of holding up colleagues. Areas that they are familiar with and areas that are easier to grasp-can be covered more quickly or in some cases omitted completely. This leaves more time available to devote to the more complex areas and means that the training can be more flexible.



1.1.5 Produce a user guide



Produce some documentation for each procedure, which illustrates step by step how each procedure is carried out using the new system. This can be in the form of a user guide.

When you come to train users, the training session will be fairly intensive and it is unlikely they will remember everything that you cover. Having a user guide to refer back to at a later date is essential. However, written instructions are no substitute for actual "hands-on" training.

1.1.6 Training evaluation and follow up

Evaluating your training program after it has taken place gives an opportunity to assess whether it has achieved the goals outlined. It also provides valuable feedback for use in future training programs. There may be areas of operating the system on which users are still unclear. These areas need to be covered in any follow-up training sessions and concentrated on more in any future training programs. Follow up training, after users have had the opportunity to work with the system for a period of time, is valuable in that it can help you to assess the ease with which certain concepts can be grasped as well as the impact of the initial training program.

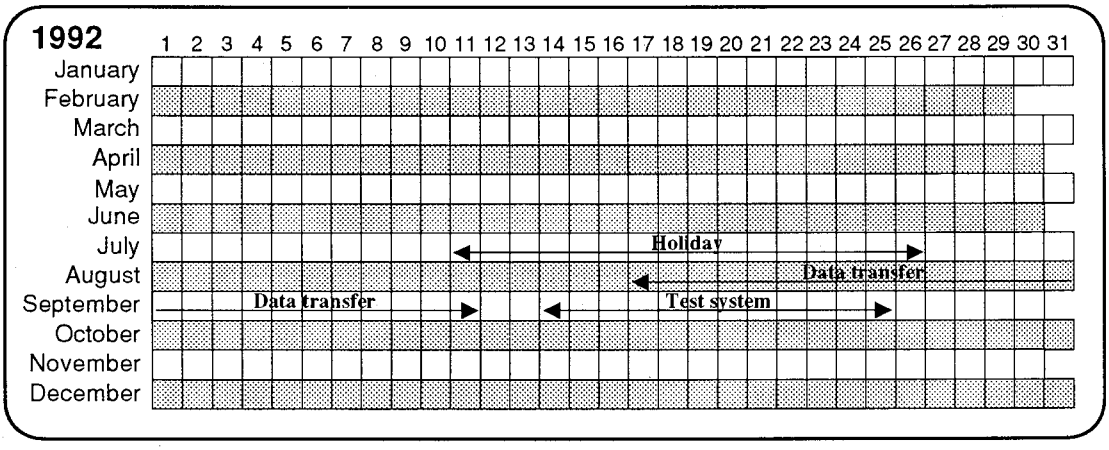
Training sessions provide an opportunity to get feedback from users on how easy the system is to use. They might have ideas on how the system could be improved, which could be incorporated in a future version of the system.

1.2 Introducing the new system

You can introduce your new system when:

- The equipment is fully functional and available for use
- All data have been transferred from the old system
- The system has been thoroughly tested
- Users have been fully trained

When a new system is introduced, there are certain to be initial difficulties and queries from the users. It is essential that you are available on the date of change over and for several days afterwards and that your time is free to oversee this change-over and deal with any problems or queries that arise.

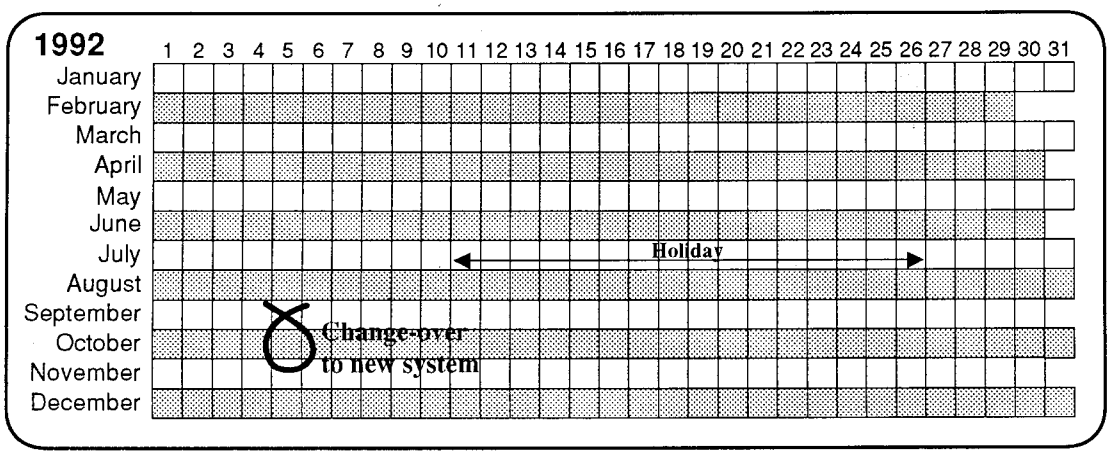


Whether you are changing to a computerized system or a new manual system, you will need to think about how the data will be transferred from your current system (if one exists) into the new system. If you have a lot of existing data, keying it into the computer or transferring it onto different manual forms can be a very time-consuming process, but you obviously cannot begin to work with the new system until all the relevant data are in place. Be realistic about the time this will take when setting deadlines for introducing the new system.

Change-over should be carefully planned

If no documentation system is currently being used, introduction of the new system can be instant. However, if an existing manual or computerized system is being used, the change-over should be carefully planned. It might need to be a more gradual process. In the following section we will look at the relative merits of the different ways of introducing the new system.

1.2.1 Immediate change-over



Immediate change-over is a realistic option only if there will be a few users of the system.

Before the change-over takes place, the system must be checked to ensure that all the data are accurate and compatible. Each procedure should also be thoroughly checked to ensure that it can be performed smoothly and accurately.

Then on a predefined date, which all users must be made aware of well in advance, the new system can be introduced and the old system discontinued. Training of all users must be carried out before this date.

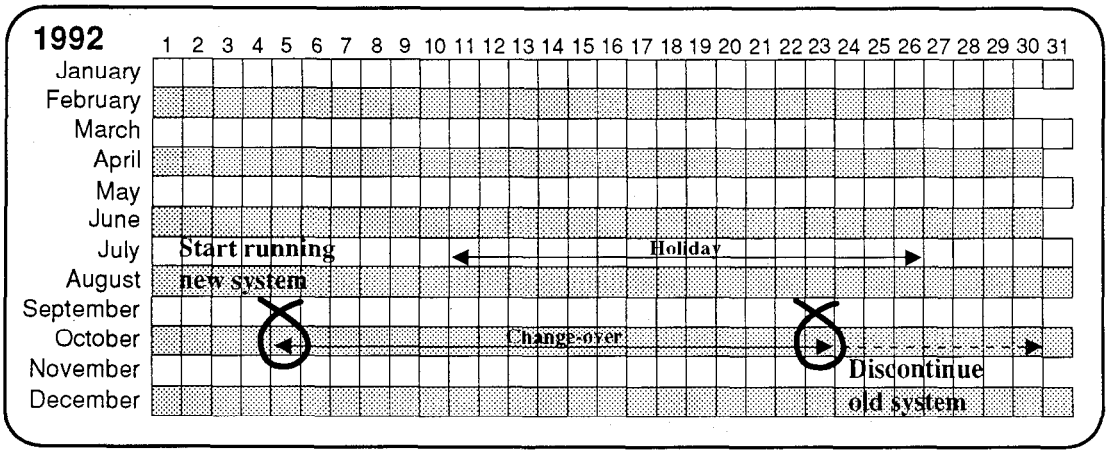
This is a simple approach, but it can run into problems if there are errors in the new system which were not detected previously. It is therefore essential that you test your system as thoroughly as possible before the date of change-over.

1.2.2 Operate the new system in parallel with the old system

It is essential that all users understand the importance of keeping both systems up-to-date

If you decide to introduce your new documentation system by operating it in parallel with the old system, it is essential that all users understand the importance of keeping both systems up-to-date. The concept behind this approach is that if there are any problems with, or errors in the new system, the old system can still be referred to for all documentation and information requirements. If both systems are not diligently kept up to date, the whole exercise is pointless. For example, if you find different values stored in the different systems, how do you know which data are correct? If you decide to introduce your system in this way, this concept should be explained as part of the training program.

With this approach, again it is essential to test your system thoroughly before change-over However, any problems with the new system will not be quite as devastating as with immediate change-over, as the old system can be referenced.



When you are satisfied that the new system is operating efficiently and accurately, the old system can be discontinued. Ensure users are informed when this happens and check that they have not encountered any problems with the new system that you are unaware of.

This can be a safe way of changing over to a new system. The main problem with this approach is that there is a large duplication of effort in running two parallel systems and confusion and mistakes are possible if both systems are not kept diligently up to date.

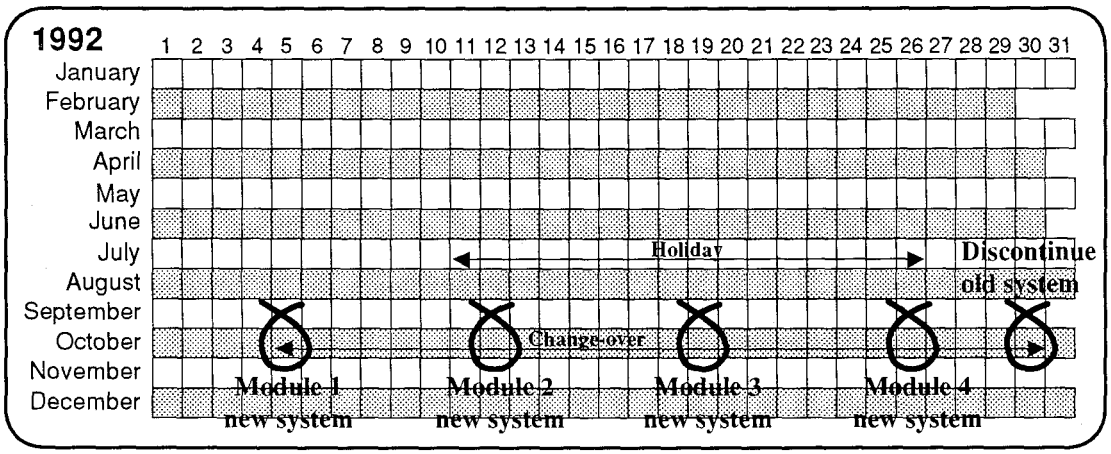
1.2.3 Phase in the new system

In this approach, the introduction of the new system is a gradual process. Rather than replacing the entire system all at once, documentation procedures are split down into separate modules and, after thorough testing, introduced one at a time. The separate modules could be phased in using either of the two previously described approaches, i.e. immediate change-over, or by operating the new procedure module in parallel with the old procedure (and using the same considerations for discontinuing the old module as set out in section 1.2.2).

The order in which the modules are introduced should be logical and **NOT** randomly chosen

The order in which the modules are introduced should be logical and not randomly chosen. So, in your genebank, a possible logical order for implementing the separate modules might be first implementing the "registration" procedure, then the "seed testing" procedure, then the "inventory" procedure and so on. This is one possible logical order. A different order (even the reverse order) might be more logical for your genebank.

The advantage with this approach, compared with the previous two approaches, is that there is an opportunity to learn from mistakes at each stage, rather than be faced with a weight of problems all at once. You should still make sure that this sort of change-over is carefully administered to avoid any confusion.



2 Data security

If a documentation system is going to be useful, easy maintenance of reliable and accurate data is essential. This is sometimes referred to as *data integrity*. Maintenance of data integrity is a vital part of the maintenance of your documentation system; In this section, we will look at ways in which data integrity can be threatened and steps that can be taken to counteract this.

There are four main ways in which data integrity can be threatened:

- *Human error* – data are entered wrongly or accidentally deleted or modified
- *Software errors* – errors in the software (known as *bugs)* causing data corruption
- *Hardware errors* – failure of hardware causing data corruption
- *Malicious damage* – by other persons or computer viruses

Maintaining data integrity involves maintaining the *security* of the system. The following sections suggest some useful practices which help to achieve this.

2.1 Procedural controls

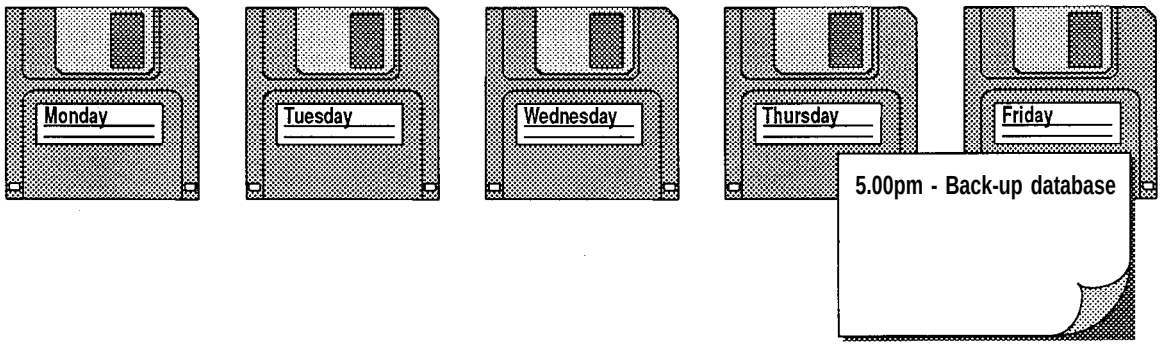
One effective way of ensuring data security is to restrict the number of users who have access to the system. This can be by simple physical restrictions (e.g. the computer or manual files are in a room which can be locked) or by using passwords to protect the computerised system (or parts of the system) from users who have not been trained in their use.

Other considerations concern using features of the software to prevent errors getting into the system. These considerations have been discussed in detail in Chapter 8. For instance, masks should be used extensively for data entry and modification to avoid accidental entry or modification of data unrelated to a particular procedure. Also fields should be set up to accept only valid responses. Importantly, users should only be able to access the system through the menus: they should not use the "back door" to the database via the database management software.

2.2 Perform regular back-ups of data files

Disasters can happen, even in the best regulated of systems. You might accidentally delete data files, your hard disk might fail, there might be a software fault which corrupts your data, a virus might infect your system, there might be a power-cut, lightning might strike the building

and destroy electrical equipment ... These situations might seem unlikely but they do happen and they *have* happened to people working in genetic resources documentation!



By making regular copies of all your important data files you will protect yourself against all possible disasters. Do this on a daily or weekly basis. Copy your data files to floppy disks and look into the possibility of storing them at a separate location. Ensure that the main user menu contains the option to back-up data files, so that becomes a routine operation. Avoid the situation where you might say "If only I had made a back-up of the data ..."

2.3 Avoiding and eradicating virus infection

In Chapter 7 we looked briefly at how your system can become infected with computer viruses. The risk of infection by computer viruses can be substantially reduced by taking some basic precautions. However, it is impossible to completely avoid infection.

If you think you might have a virus, you should first switch off your computer. Viruses cannot do any damage when the computer is off. In the following sections we will look at these precautions and also study the steps you can take to rid your system of viruses should you be unlucky enough to become infected.

2.3.1 Do not copy software illegally

Copying software is a common source of infection, as if the software you are copying is infected, your system will become infected too. For much software, it is also illegal. Therefore, only install software that you have bought from a commercial supplier as these are generally examined for viruses before distribution. Make sure that the disk is sealed in its original packaging, or if a sales-person installs the software on your computer, make sure they use the manufacturer's original copy not their own copy. The sales-person's copy might have been used on other computers and could be infected with viruses.

2.3.2 Use anti-virus software

There are many commercially available software packages which you can install on your system to guard against and eradicate virus infection. Some anti-virus software is automatically activated when your computer starts up, and will routinely check all disks you use. Other anti-virus software requires you to instruct it to check any new disks you are using. Certain software can create a copy of vital areas of your disk which will help you recover from potential disaster. However, you need to create this copy *before* any damage occurs. Study your anti-virus software manual on how to do this.

Anti-virus software can only guard against viruses it has been designed to recognise. As there are new viruses appearing all the time, the software won't be able to stop all possible infection, but it can significantly reduce the risk.

The back-up you created of your system is indispensable if your system becomes infected by a virus

If viruses cannot be removed by anti-virus software, it may be necessary to erase all the infected files and re-install them from the original disks, but this should only be a last resort. For boot sector viruses, you may also have to reformat the entire disk. The back-up you created of your system is indispensable if this happens.

If you are unsure how to remove viruses and prevent them from re-infecting the system, you should seek professional advice.

2.3.3 Write-protect your program disks

When installing new commercial software from floppy disks, *write-protect* the disk before you install the software. This involves placing a write-protect tab on the disk. Consult your computer handbook to find out how to do this (see also figs. 3 and 4). A write-protected floppy disk cannot be infected by computer viruses. However, with some software, part of the installation process involves writing to the floppy disk and so write-protecting the program disk would not be an option.

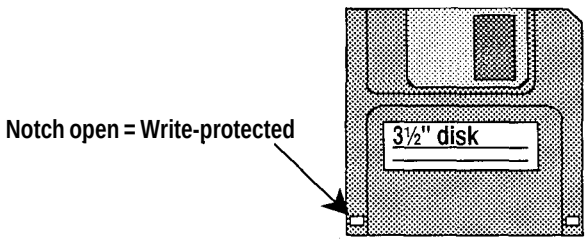


Fig. 3. Write-protected 3½ inch disk

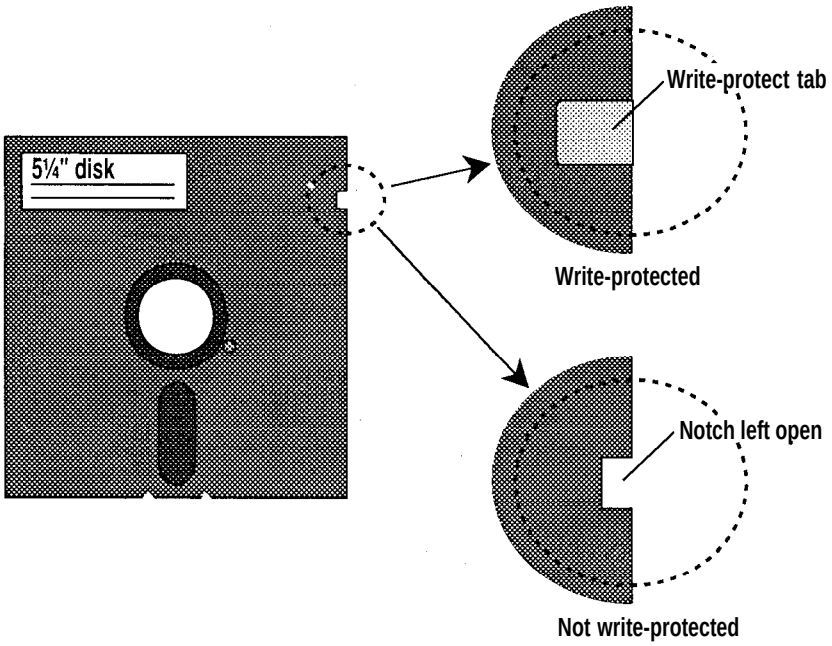


Fig. 4. How to write-protect a 5¼ inch disk

2.3.4 Keep a virus-free, write-protected system floppy disk

Under normal circumstances, your system will reference the hard disk when starting up. As we mentioned in Chapter 7, it is possible to start your system from a bootable floppy disk, known as a *system disk.* It is good practice to keep a write-protected system disk in case your hard disk becomes infected with a virus.

2.3.5 Make program files "read-only"

Consult your computer handbook to find out how to make your program files "read-only". This won't prevent all virus attack but it's a simple preventative measure that reduces the risk of infection.

Remember: *prevention is better than cure.* Adopt basic safety standards and make sure that all users of the system know what these standards are.

3 Data exchange between documentation systems

The ability to exchange data between your documentation system and databases on separate computers or sites or between different applications on the same computer is a very important consideration. If data can be exchanged easily, the value of your documentation system will be considerably enhanced. Occasions when you might want to do this include:

- Exporting/importing data to/from other database systems
- Exporting data to a word processor or desktop publishing software for the preparation of high quality reports or catalogues
- Exporting data to a spreadsheet or to statistical software in order to perform complex calculations
- Import of data from spreadsheet or statistical software. (This allows you to perform calculations and analyses on your data before it is stored in the documentation system).

It helps considerably if databases at different sites are managed by the same software and if database files have the same structure and use the same formats for the data

You may wish to use data from a database at another genebank and similarly, they may wish to use data stored in your database. If you're going to be exchanging data between databases on different sites on a regular basis, you'll want it to be a smooth and painless procedure. It helps considerably if the databases are managed by the same software. It helps even more if the database files at different sites have the same structure and use the same formats for the data.

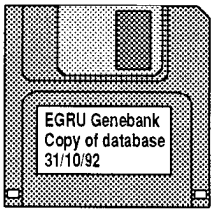
Aim for uniformity across the sites

If you are setting up a documentation system that will operate on more than one site (e.g. at different research stations) you should aim for uniformity across the sites. In practice, this means that all the sites will operate the same version of the menu driven documentation system you use yourself, perhaps using fewer documentation procedures. This will make data exchange more straightforward.

Remember to make sure that data on your genebank accessions can be distinguished from data on accessions at other stations stored in your database. If different genebanks operate different accession numbering systems, the accession number will identify where the accession is stored.

3.1 Methods of data exchange

Sometimes, data exchange with other computers simply involves copying files from your computer onto a floppy disk and then retrieving files from the floppy disk on another computer. Floppy disks are a convenient method of exchanging data between computers. They are portable and easy to send by mail to contacts at other genebanks.



Data can also be exchanged between computers which are *networked.* A computer network is a group of computers, usually in one building or on a single site which are-physically linked together using cables.

If computers are situated in different geographical locations, it is possible to exchange data files using national and international telecommunications systems and the appropriate communications software.

3.2 Exchange selected data

If you plan to exchange data regularly, explore the features of your software to see if it is possible to "flag" data that has been added or modified since data was last exchanged. Some database management software enables you to do this by giving you the option to store the date when a record was last modified and the date a new record was added. Data modified after the last date of exchange can then be selected and transferred, rather than transferring a copy of the entire database.

3.3 Problems in data exchange

Unfortunately, data exchange is not always straightforward. In the following sections, we will try to look at some of the problems encountered when exchanging data.

3.3.1 Different descriptors or descriptor states used

Data exchange between databases is usually straightforward if files are identical in structure, i.e. they have the same field structure, field names and field specifications, the same quantitative and qualitative scales and the same systems of coding.

If the fields do not have the same specifications, you risk losing or corrupting your data in transfer.

Table 1 lists some examples of where data exchange between two files is problematic because of the different conventions used.

Table 1. Examples of where data exchange between two files is problematic

DESCRIPTOR	GENEBANK 1	GENEBANK 2	COMMENTS
Collection source	2 = farm land	2 = backyard	Different codes used for the nominal scale
Mature leaf size	9.5 cm	5	Different descriptor definitions as a result of the use of different types of scale: continuous and ordinal
Monthly rainfall	17mm	1.7 cm	Different descriptor definitions as a result of the use of different units
Photograph taken	+	Y	Different codes used for binary scale

Try to ensure that the **SAME DESCRIPTORS** with the **SAME DESCRIPTOR STATES** are used in different databases if you wish to exchange data between them

You should try to ensure that the *same descriptors* with the *same descriptor states* are used in different databases if you wish to exchange data between them. It is not impossible to exchange data when descriptors and descriptor states are different, but the process is more complicated. For example, if a descriptor has been recorded in separate files using different scales or a different set of codes, before you can exchange data; it needs to be translated to the system of scales or codes used by the recipient file.

Note: If your database requires you to specify field widths when defining fields, make sure these are identical for the fields you are exchanging data between. If they are different, you run the risk of losing data.

If there are people with computing expertise at other sites with whom you intend to exchange data, you should work with them to make sure that there is uniformity in descriptors used and compatibility in any file structures that they develop.

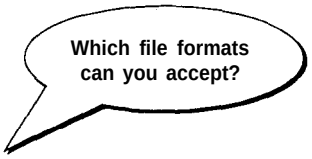
3.3.2 Different file formats used by different applications

If data are being exchanged between different types of software packages, the process is more complicated. The format of a file will differ depending on the software used to generate it. Different software packages for database management, spreadsheets, word processing, desktop publishing, graphic design and statistical analysis all use different file formats, which are not directly compatible.

However, most commercial software packages have data import and export facilities, for converting to and from file formats other than the one used by the program itself. Also, the facility to convert to certain standard file formats is usually available.

ASCII format is a commonly accepted file format

The ASCII file format (American Standard Code for Information Interchange) is probably the mostly widely accepted text-based file format. Most commercial software packages such as database managers, spreadsheets, statistical analysis packages, word processors and desktop publishers can produce and process ASCII files.



If you are exchanging data with another genebank, check with them the formats in which they can accept data. Also inform them of your preferred file format. If in doubt, use the most commonly accepted file format: for database management software. this is often fixed field ASCII.

3.3.3 Different operating system

What operating system do you run on your computer?

Floppy disks were mentioned as a convenient and simple method of exchanging data between computers. However, when exchanging data between computers running different operating systems, the process is not quite so straightforward. In order to exchange data on a floppy disk, conversion software must be used. This will enable your computer to read data from a disk generated by another computer running a different operating system and vice versa.

is your computer linked to a network?

Alternatively, data can be exchanged over a computer network, as computers with totally different operating systems can be linked together on a network. This provides a useful route for data exchange between computers running different operating systems.

3.4 Points to bear in mind when exchanging data

Whether you are exchanging data between databases managed by the same software, databases managed by different software or between totally different types of software packages using different file formats, there are a number of points you need to clarify before attempting data exchange. These are illustrated in the flow chart in fig. 5.

NEVER use existing data in the test procedure

You should always check thoroughly the accuracy of data exchange with test data- *never* use existing data-in the test procedure as you might lose valuable data in the process. Also, as a safety measure, always make a back-up copy of the different files involved *before* you exchange any data. That way if things go wrong, you will always be able to recover any data that get corrupted or lost in the process.

Make a back-up copy of the different files involved **BEFORE** you transfer any data

3.5 Data exchange between spreadsheets and databases

Spreadsheets generally have a greater facility for statistical analysis than database management software and it is for this reason that they are commonly used in genetic resources work. In order to use spreadsheets in conjunction with database management software, a system must be developed for straightforward and error free data exchange which is not labour intensive.

Data exchange between spreadsheets and databases often involves re-organisation of data before it can be exchanged. For example:

- The file structure may need to be modified so that the recipient software package can recognise the different data elements
- The data themselves may need to be transformed to comply with descriptor states, scales and codes used
- The field specifications may have to be modified to match the specifications defined in the recipient software package

It is worth spending time devising the most efficient method of data exchange, especially if you need to exchange the same files on a regular basis. Consult your manuals for the best way of doing this. Use the data management features of your software to set up a routine option for regular data exchange and include this as an option on your menu system.

It is often much faster and less complicated to transfer data from a database to a spreadsheet than to transfer data in the other direction. Look at your requirements for data analysis and ask yourself whether your data can first be recorded in your database and then exported to a spreadsheet for analysis.

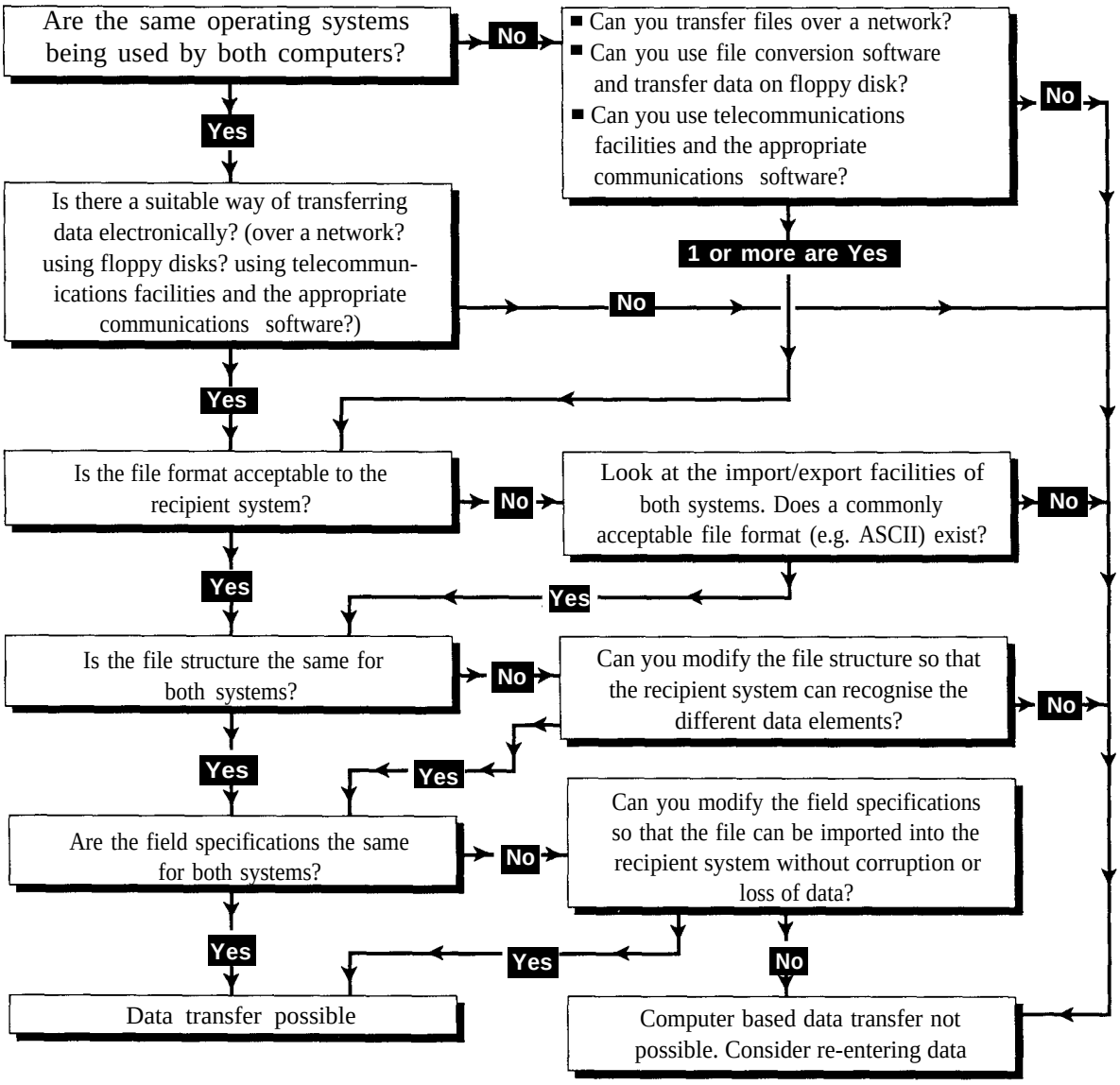


Fig. 5. Considerations for computer based data transfer

4 Modifying your documentation system

You have designed a documentation system which satisfies the genebank's current documentation and information needs and future needs have been anticipated as far as possible. So, after spending a lot of time and effort designing and implementing the system, the very thought of change might be an alarming one.

4.1 Reasons for change

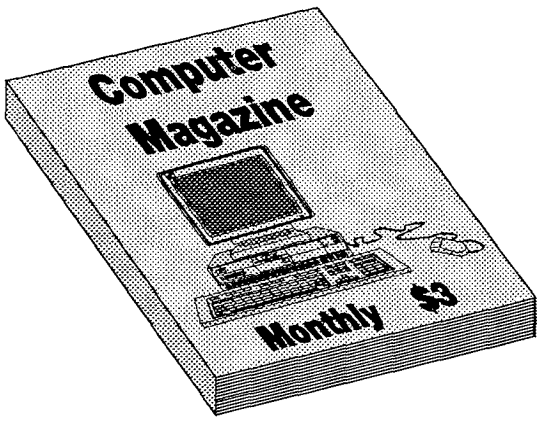
Genebank documentation systems are rarely static and unchanging. Instead they are dynamic and evolve to accommodate the changing documentation and information needs of the genebank.

4.1.1 Changes in genebank objectives

Genebanks tend to expand over a period of years. As they do, policies change, new crop accessions are accepted, new areas of work undertaken, current areas of work are phased out, new collaborations with other institutes are entered into. All these changes will have an effect on the information requirements of the genebank and on the operation of the documentation system.

For example, as the areas of work change, new documentation procedures will need to be developed and existing procedures modified or phased out. It may be necessary to develop new reports for the changing information needs. New screen forms may need to be designed. Menus may have to be altered to include new procedures or exclude redundant procedures. New data files may need to be created and integrated into the computerized system.

4.1.2 Changes in information and documentation technology



Exciting developments in the capabilities of new microcomputers are continually taking place. For example, increased speed of operation, increased storage capacity, more sophisticated software and better facilities for communicating with other computers are to name but a few. These technological advances are becoming continually more affordable and will have a big impact on the way that your documentation system can operate. Try to keep yourself aware of any developments in technology that would improve the efficiency of your system.

Take advantage of the software updates. Software is updated with the needs of the user in mind. So assuming you have chosen a suitable package to start with, any updates should help improve the efficiency of your system.

4.1.3 Changes in staff

It's unlikely that the same people will be working in your genebank in, say, five years time. People change jobs, new staff are employed. Staff changes inevitably bring about changes in the documentation and information requirements. For example, if a new curator is appointed, would he or she have the same information requirements as the outgoing curator? If a new seed handling officer is appointed, he or she might have ideas on how to streamline procedures or organise work more efficiently. Any such changes will have an effect on your documentation procedures and the information demands on the system.

4.2 Reviewing the system

Periodically, you will need to review your documentation system in order to:

- Determine whether it is satisfying current documentation and information needs
- Identify areas which can be improved
- Identify new features for incorporation

This review of your documentation system should be approached in a different way from the genebank analysis you performed prior to designing the system. The questions you need to ask are different. Let's look at some of the issues you need to consider in such a review.

4.2.1 Genebank objectives

Have the genebank objectives changed since the documentation system was set up? If they are different, how do they affect the documentation and information requirements (if at all)? Does your documentation system still satisfy the information requirements?

4.2.2 Operation of the system

Users should be consulted for feedback

Is the system user-friendly? Users should be consulted for feedback on whether any operations need to be modified or manual and screen forms redesigned. Also there may have been changes in staff since you designed the system and the new staff may have different needs from the previous staff.

4.2.3 Efficiency of the system

Is the system making the best use of available resources (e.g. computers, staff)? It may be that certain documentation procedures you have designed are inefficient or impractical. For instance, data entry procedures for a computerized system might be impractical (because the computer is located in an inconvenient place) or inefficient (slow or involving an undesirable duplication of effort).

If you are currently using a manual documentation system, could your data be managed more efficiently using a computerized system? If the volume of data being documented is constantly increasing, this is an option to consider.

If you are using a computerized documentation system, would new hardware or software significantly increase the efficiency of the system?

4.2.4 Accuracy of the system

Are all the documented data reliable, accurate and up to date? Is it easy to detect and correct errors in the system? Are the reports useful?

Try to identify how incorrect data are entering the system and the necessary steps to take to counteract this. Use the data management features of your software to disallow invalid data from entering the system.

4.2.5 Additional information requirements

Are all the reports that you are routinely generating useful? Are any other reports required?

Talk to the people who use these reports or requested that they be generated, to get their comments on these points.

4.3 Planning changes in the system

After reviewing your system, you will have a good idea whether it is operating satisfactorily and will have identified -areas which require modification. If you are working in a small genebank and are solely responsible for the documentation; this review-will not take very long; you will already be aware of any problem areas!

You should not modify the system without carefully thinking through the consequences. All modifications must be carefully planned

However, it is important to demonstrate a need for modifying the system, even if the suggested modifications are minor ones. Above all, you should not modify the system without carefully thinking through the consequences. Otherwise, you might accidentally introduce errors which can affect documentation procedures and information retrieval.

This is particularly true if you modify the structure of data files. Changes in field specifications will affect the operation of screen entry forms, the appearance of reports and the ability to exchange data.

Any modifications to the documentation system need to be planned. This involves a number of steps which are illustrated in the chart in fig. 6.

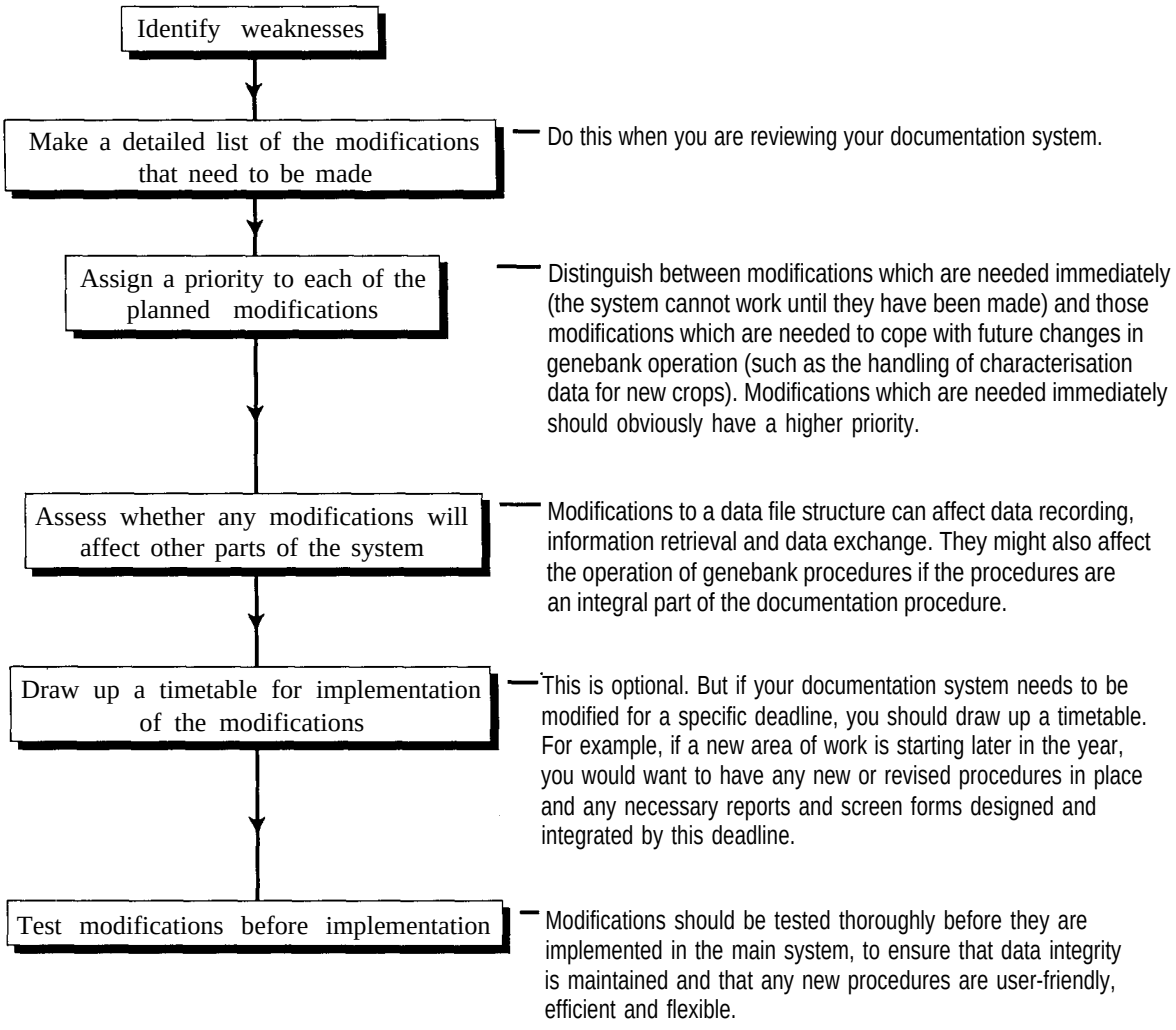


Fig. 6. The planning stages for the modification of the documentation system

4.4 Importance of a flexible documentation system

Genebanks are dynamic bodies. They are constantly evolving to support the global conservation effort and to meet the needs of researchers in the field of plant genetic resources. The genebank documentation system plays a central role in the support of all genebank activities and the planning of future projects. Periodic modification of the documentation system is therefore inevitable. Flexible design of your documentation system means that any such modifications will be much easier to implement.

5 Where to go from here

After working thoroughly through Chapters 1 to 10 of this guidebook, you will have gained the necessary knowledge to design and implement a new documentation system for your genebank.

It is important to give a lot of thought to the best way of introducing your new system. Ensure that you are available on the day the system is implemented and for several days afterwards to deal with any problems and queries that crop up.

Training should be an ongoing process. As staff leave and new staff join, the new staff will require training in the use of the system. As new computer equipment is added to your system and software updates are implemented, users will require training in the use of this new equipment and software. Modifications will have to be made to your basic training program, to incorporate these changes.

You should try to keep up to date with developments in the area of data exchange and ways in which you can improve the data exchange facilities of your system. Efficient data exchange facilities will greatly increase the value of your documentation system.

Do not become neglectful of data security. Ensure that rigorous procedures are in place to maintain data security and that all users are aware of the importance of following them. This should be emphasised in all training programs.

A carefully designed documentation system which is operated in accordance with the recommendations given in this guidebook, will provide a valuable resource for you and your colleagues at the genebank.

6 EXERCISES

Exercises

1. Why is it necessary to have a system manager? List some of the duties that such a person might perform.
2. List the three ways of introducing a new documentation system. What are the advantages and disadvantages of each approach?
3. Adequate training in the operation of a documentation system is essential to the success of the system. State briefly the considerations which have to be taken when constructing a training programme with regard to the available staff resources and their training needs.
4. What steps would you take to reduce the risk of virus infection in your system?
5. List ways in which data exchange between documentation systems can be facilitated. What precautions should you take when developing procedures for exchanging data?

Glossary

A

accession

A distinct sample of *germplasm* which is maintained in a *genebank* for *conservation* and use.

accession number

A unique identifier assigned to each *accession* when it is registered with a *germbank*.

active collection

A *collection* of *germplasm* used for *regeneration, multiplication,* distribution, *characterisation* and *evaluation.* Ideally germplasm in the active collection should be maintained in sufficient quantity to be available on request. Active collection germplasm is commonly duplicated in a *base collection* and is often stored under medium to long-term storage conditions.

allele

An alternative form of a *gene* which can have different properties when expressed.

ASCII

American National Standard Code for Information Interchange. A widely used standard for data exchange.

B

back-up

A secure system, device, or *file* which can be used in the event of *hardware* problems or *data* loss/ corruption.

bar code

A pattern of light and dark lines on labels which can be read by a light pen for direct entry into a computer.

base collection

A *collection* of *germplasm* which is kept for long-term, secure conservation and is not to be used as a routine distribution source. Seed is usually stored at sub-zero temperatures and low moisture content.

batch

Germplasm from a specific *regeneration* or *multiplication* cycle of an *accession.*

batch reference

Any identifier such as a date, code or number which uniquely distinguishes the *batch* of a particular *accession.*

binary scale

A scale for scoring *data* where there are only two possible responses.

boot sector

The area of a *disk* which is accessed each time the computer starts up for instructions on how the computer will operate.

C

callus

A group of dividing, undifferentiated plant cells formed on or below a wounded surface. It is the first stage in the *regeneration* of such tissue. Callus cultures can be produced *in vitro* by treating an excised fragment of tissue with a growth regulator.

CD-ROM

Compact disk read only memory. An optical *disk* which can store large amounts of *information* such as literature *databases* and libraries of computer graphics. CD-ROMs require a specialised *drive* called a CD-ROM player in order to read the disks. As the name suggests, the disks allow reading of the stored *data* but no writing or editing of data. See also *WORM*.

central processing unit
The main *microprocessor* in a computer which works with and controls other components. Often called the CPU.

CGA
Colour graphics adapter. See *monitor*.

character
1. A recognisable quality or attribute resulting from the interaction of a *gene* or group of genes with the environment.
2. A letter, digit, punctuation mark or special graphic used for the production of text.

characterisation
The recording of those *descriptors* which are highly heritable, can be seen easily by the eye and are expressed in all environments. See also *evaluation.*

chip
See *silicon chips.*

clone
A group of organisms or cells descended from a common ancestor that are genetically identical. They can be produced by apomixis, parthenogenesis, vegetative propagation or tissue culture.

collection
A group of *germplasrn accessions* maintained for a specific purpose under defined conditions. See also *base-, active-, working-, field-* and *in vitro-.*

colour graphics adapter
See *monitor.*

conservation
The management, preservation and use of known *genetic resources* so that they might yield the greatest sustainable benefit to the present generation while maintaining their potential to meet the needs of generations to come.

continuous scale
A scale for scoring *quantitative data* for which the number of potential values is not pre-defined and potentially limitless. For instance, the *descriptor* "seed weight" measured in grams.

CPU
See *central processing unit.*

crop
A *species* expressly cultivated for use.

cryopreservation
The preservation or storage in very low temperatures, usually in liquid nitrogen (-196°C).

cultivar
A *crop* variety produced by scientific breeding (as in modern, improved or high yielding variety) or farmer's selection methods (as in primitive cultivar or *landrace).*

cursor
1. A flashing symbol on a screen indicating where the next *character* can be entered.
2. A symbol, such as an arrow, which is displayed on a screen and moved with the aid of a *mouse* (or special cursor keys).

D

daisy wheel printer
A basic type of printer which generates characters by striking an inked ribbon with a moulded character.

data
Quantitative or *qualitative* values derived from observations.

database
1. General: An organised set of *data* assembled for a specific purpose and held in one or more storage *media.*
2. Computing: A well-organised set of inter-related *data* held in one or more *files* which are capable of being managed by the same *software.*

database management system software
Software specifically developed to manage *data* stored on a computer. Most database management system software allows you to perform the following basic activities: data entry, modification and deletion, *information* retrieval and data exchange.

data dictionary
A description of the meaning, relationship, usage and format of *data* within the data structure of a *database.*

data export
The extraction of selected *data* from a computer application.

data import
The incorporation of *data* originating from another computer application.

data integrity
The reliability and accuracy of *data*

data logger
Commonly a portable, hand-held electronic device used for recording *data* which can later be transferred to a computerised system.

data redundancy
The unnecessary duplication of *data* or *information* in a *documentation system.*

DBMS
See *database management system software.*

descriptor
An identifiable and measurable characteristic used to facilitate *data* classification, storage, retrieval and use.

descriptor list
A collation of all the individual *descriptors* used for a particular *crop* or *species.*

descriptor state
A clearly defined state which a *descriptor* can take.

digitisation
The process of computerizing *data* from hard copy. See also *graphics tablet.*

directory
A unit of organisation On a *disk* in which *files* can be grouped together.

disk
A device used for data storage. See also *hard disk, floppy disk* and CD-ROM.

disk drive
A device used to read and write *data* from and to *disks.*

DNA
Deoxyribonucleic acid. It forms the molecular basis for heredity and is principally found in chromosomes and also in mitochondria and chloroplasts.

documentation system
Any way of storing and maintaining *data.* A documentation system can use manual methods (such as hand written *records)* and/or completely computerised methods for data storage and maintenance. The system is also designed. for *information* retrieval.

dominant
Descriptive of an *allele* which is expressed regardless of the nature of another allele of the same gene.

donor
In *passport data:* the institution or individual responsible for donating the *germplasm.*

DOS
Disk operating system. See *operating system*

dot matrix printer
A basic type of printer which generates characters by electronically manipulating a matrix of tiny needles to form the required characters.

E

EGA
Enhanced graphics adapter. See *monitor.*

embryo
The rudimentary plant within a seed which arises from the zygote or sometimes from an unfertilized egg cell.

enhanced graphics adapter
See *monitor.*

evaluation
The recording of those *descriptors* whose expression is often influenced by environmental factors. See also *characterisation.*

F

fax modem
A specialised *modem* which can be used to send and receive faxes.

field
In a *record,* it is a specified area which always holds *data* for a particular *descriptor.*

field collection
A *collection of germplasm* maintained as living plants (e.g. fruit trees, glass house crops and field crops). *Germplasm* which would otherwise be difficult to maintain as seed is commonly maintained in field collections. It is sometimes the only type of collection maintained by a *genebank.*

field name
The label given to a particular *field.*

field width
The number of *characters* that a *field* can accommodate.

file
A set of *data* (or a *program)* which is treated by the *operating systern* as a single unit.

flag
A clearly visible mark to indicate a particular condition.

flat file manager
A simple *database management system* which works with one *file* at a time. More sophisticated flat file managers can handle several files at a time and resemble *relational database managers.*

floppy disk
A portable computer *disk* with a low storage capacity. Two sizes are commonly used – 3½" and 5¼". Floppy disks are commonly used for putting new *programs* or *data* onto the *hard disk* or transferring programs/data between different computers.

font
A particular typeface and size including all the characters produced (uppercase and lowercase letters, numbers and punctuation marks). Fonts can either be monospaced (each character takes up the same space) or proportionally spaced (the space occupied varies from character to character.

forage species
A plant *species* which is eaten by animals unharvested.

format
The arrangement of *data* in a *file.* With *database management software,* it can refer to the *record definition* used.

G

gene
A basic unit of inheritance which is encoded in a sequence of DNA.

genebank
A *genetic resouces* centre where *germplasm* is maintained in one or more *collections.*

genetic resources
Germplasm of plants, animals or other organisms containing useful *characters* of actual or potential value.

genus
A taxonomic grouping of similar *species.*

germination test
A procedure to determine the proportion of seeds that are capable of germinating under particular conditions.

germplasm
The genetic material responsible for a plant's characteristics.

graphics tablet
A flat plate on which a paper containing an image is placed. Using a special pen, a series of positions on the paper can be marked and these coordinates are entered directly into the computer. This process is commonly called *digitisation.* This has great application when dealing with data from maps. Also called digitiser, digitising table.

group data
Data concerning groups of *accessions.*

H

hard disk
A computer *disk* with a large storage capacity which is usually located inside the computer and is often not removable.

hardware
The name given to all the physical items which comprise a computer.

heterogeneity
Synonymous with *variation.*

I

identifying field
A *field* which is used to select one or more desired *records.* One or more identifying fields in different *files* can be used to form a *relationship.*

***in vitro* collection**
A *collection* of *germplasm* maintained as plant tissue grown in active culture on solid or in liquid medium. Germplasm can be maintained as plant tissue ranging from protoplast and cell suspensions to *callus* cultures, *meristems*/shoot-tips and *embryos.* In some cases the tissue is stored at very low temperatures such as under liquid nitrogen (so-called *cryopreservation)*

index
Usually a separate *file* which stores *information* on the location of specific *records* in a *database* file based on a single *field* or combination of fields.

information
The meaning that emerges when *data* are recorded, classified, organised, related or interpreted.

ink jet printer
A "non-impact" type of printer. Electronic commands instruct ink drops to be squirted onto the paper to form the characters.

institutional genebank
A *genebank* set up to conserve only the *germplasm* which is used in (or is potentially useful to) the research programmes in its host institute or host agricultural research centre.

interface
Software or electrical equipment where interaction occurs between two systems. For instance, an *operating system* is an interface between the *hardware* and the user of the system.

inventory
A detailed list of *accessions* held in a *genebank* usually with regard to their maintenance and storage.

isozyme
Multiple forms of an enzyme in an organism which can be distinguished by electrophoresis. Also called isoenzyme.

L

landrace
A set of populations or clones of a *crop species* produced and maintained by farmers.

laser printer
A printer which uses laser technology to produce high quality images for both text and graphics.

layout
The basic design of a page.

logical system design
A model *documentation system* based on user requirements, outlining how the system might operate regardless of *hardware* and *software* considerations. See also *physical system design.*

M

mandate crop
A *crop,* responsibility for whose study and conservation has been undertaken by a particular organization.

mask
A sub-set of the total number of *fields* available in a *view.*

math coprocessor
A *microprocessor* used to process mathematical calculations.

mean
A measure of the central tendency of a set of numbers obtained by dividing the sum of the numbers by the quantity of numbers.

media
Any material or electronic device used for *data* recording *e.g.* paper forms, magnetic tape.

memory
The computer's electronic working space. The size of the memory affects the size of *program* a computer can handle and the speed a program can run. See also *RAM, ROM.*

menu
A display of the available options for selection by the user of the system.

meristem
A localized region of rapidly dividing, undifferentiated cells from which new cells arise that differentiate into specialized tissues. Meristems are found in growth areas, e.g. shoot and root tips.

microchips
See *silicon chips.*

microcomputer
A type of computer commonly found in the laboratory, office or home that can be used on a desk. Synonymous with personal computer.

microprocessor
A customised and powerful *silicon chip* which works with and controls other components of a given system. The main microprocessor in a *microcomputer* is often called the *central processing unit* or CPU.

modem
A piece of equipment which converts signals generated by a computer into a form suitable for

transmission over a telephone line. A modem at the receiving end converts the signals back for use by the receiving computer. A modem therefore allows communication between computers. Modems are commonly used to access remote computers which contain large *databases (e.g.* literature databases) or specialised *software (e.g.* for molecular biology). They can also be used to send and receive electronic mail.

monitor

A device rather like a television screen which is used to display what has been typed in and any output from *software* which is running. Also called a video display unit or VDU.

monitoring

The periodic checking of *accessions* for acceptable levels of *viability* and quantity of *germplasm.*

mouse

A small, hand-held plastic box with one or more buttons on top, that communicates with the system unit (or sometimes the keyboard) via cable or radio signal. When the mouse is moved on a flat surface there is a corresponding movement of a *cursor* on the screen. Instructions are given by pointing at, clicking on and dragging objects displayed on the screen.

multiplication

The growing out of a *batch* of an *accession* to increase the quantity of conserved material.

mutation

A change in the genetic material in the cell that results in an abnormal characteristic being expressed. Changes can be quantitative (the number of genes or chromosomes) or qualitative (the structure of the genetic material).

N

national genebank

A *genebank* set up as a national plant *genetic resources* centre maintaining many different *germplasm* samples of current and potential interest for people working in plant research nationally. Commonly it will contain germplasm which has been collected nationally. Also it may be closely associated with a research programme or undertake its own research. A national genebank can be a collaborative venture between national institutes or under the responsibility of one institute which collaborates with other national institutes.

network

An interconnected group of computers. At a simple level, a network can be a group of computers in a single building which are physically linked together with cables. At a more sophisticated level, a network can be computers on geographically separate sites which are communicate with each other using telecommunications systems.

nominal scale

A scale for scoring *qualitative data* using a series of pre-defined values. For instance, with the *descriptor* "flower colour", a nominal scale lists the different possible flower colours that can be observed.

normal distribution

The name given to a pattern of variation commonly observed in biological *data,* commonly portrayed by a bell-shaped curve.

O

operating system

The *interface* between the user and the computer. It controls the operation of *programs* and communication with the keyboard, *monitor, mouse, disks,* printers and any other *peripherals* attached to the computer.

ordinal scale

A scale for scoring *quantitative data* using a series of pre-defined intervals arranged in a logical sequence.

P

passport data

Information about the origin of an *accession* (such as details recorded at the collecting site) and any other relevant *information* including *descriptors* which assist in the identification of the accession.

pathogen

An organism capable of causing a disease in another organism.

PC

Personal computer. Synonymous with *microcomputer.*

pedigree

The record of the ancestry of an individual, genetic line or variety.

peripheral
Equipment that works with a computer but is not part of the computer itself such as a printer, *modem, tape streamer* and *graphics tablet.*

***personal* computer**
Synonymous with *microcomputer.*

physical system design
A detailed specification of how a *documentation system* might operate in a specific environment and with the chosen *hardware* and *software.* See also *logical system design.*

phytosanitary certificate
A certificate from a recognised plant *quarantine* service, that indicates that a sample is substantially free from diseases or pests.

program
The name given to a series of logical instructions to the computer which perform a particular task.

protoplast
The osmotically sensitive structures formed when the cell wall is completely removed, usually by enzymatic action.

Q

qualitative data
A non-quantitative description of the character(s) being examined *(e.g.* "brown", "hairy", "horizontal").

quantitative data
Numerical values derived from measurements or observations.

quarantine
The official confinement of plants subject to phytosanitary regulations for observation and research, or for further inspection and/or testing.

R

RAM
Random access memory. The part of the *memory* usually used when working with programs or files. Information contained in the RAM is lost when the computer is switched off.

random access memory
See *RAM.*

RAPD
Random amplified polymorphic DNA. A comparative study (among individuals, populations or species) of the DNA fragment lengths produced in controlled DNA synthesis reactions started with short sequences of DNA (primers).

raw data
Original experimental observations.

read only memory
See *ROM.*

read-only
A type of access to data that allows them to be read but not modified. In some cases, data which are read-only cannot be copied or printed.

recessive
Descriptive of an *allele* which is not expressed in the presence of another *dominant* allele of the same gene.

record
A set of related *fields* which are handled as a unit.

record definition
A detailed description of the structure of a record and the characteristics of each field. See also *data dictionary.*

regeneration
The growing out of a *batch* of an *accession* to restore the *viability.*

regional genebank
A *genebank* set up as a collaborative venture between a number of countries in the same geographical region to conserve the *germplasm* from that region and to support plant research.

registration
The acceptance of a new *accession* into a genebank, the allocation of a unique *accession number* and the incorporation of data accompanying the sample into the documentation system.

relational database manager
Database management software which can handle several files at a time, linking them together using *fields* which are common to the files. Unlike flat file managers, the common fields are stored once only and not duplicated in the different files.

relationship
Produced when two or more *files* are linked together using one or more common *fields.*

report

An output document prepared by *database management software* for the purposes of *information* retrieval.

RFLP

Restriction fragment length polymorphism. A comparative study (in individuals, populations or species) of the DNA fragment lengths produced by particular restriction enzymes.

ROM

Read only memory. Contains *information* which can be read but not changed. The instructions a computer needs in order to start working when the power is switched on are contained in ROM. Information contained in ROM is not lost, even when the computer is switched off.

S

scanner

A device which makes an image of a page and stores it as a computer *file*, in much the same way that a photocopier makes an image of a page and prints it. Scanners are widely used in scientific work for the analysis of electrophoretic gels. In the office, scanners together with specialised *software* are useful for reading text from a page and converting to *ASCII* characters.

scientific name

An internationally recognised designation used for the naming of organisms. Scientific names are commonly expressed binomially *(e.g. Secale cereale, Triticum durum);* or trinomially where greater precision is required *(e.g. Brassica oleracea* var. *botrytis, Brassica oleracea* var. *gemmifera).* In scientific publications it is also common for the authority to be listed *e.g. Manihot esculenta* Crantz.

screen form

A visual *interface* between the user of a *documentation system* and the *data files.*

silicon chips

Thin slices of silicon (approximately 5mm square by 1 mm thick) which contain several thousand electrical circuits. They are very powerful devices which have an enormous number of applications in the electronics industry.

software

The general name given to a *program* or group of programs which perform a particular task e.g. software for *database management,* software for word-processing, etc.

somaclonal variation

The *variation* found in *somatic* cells growing in culture.

somatic cells

All cells which are not involved in the process of sexual reproduction *e.g.* vegetative cells.

species

A group of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups. Many common names of organisms denote species *e.g.* tomato *(Lycopersicon lycopersicum),* groundnut *(Arachis hypogaea).*

spreadsheet

Software used for performing a variety of calculations. It is useful for statistical analysis and in business applications.

SQL

See *structured query language.*

standard deviation

A statistical measure of the extent of variation around the *mean* for *quantitative data.*

storage

The keeping and maintenance of *accessions* in a *genebank.* Seeds differ with respect to the conditions under which they may be stored without loss of *viability.*

structured query language (SQL)

An interactive query language which facilitates *data* operations such as searching and reporting in certain *database management software.*

subculture

1. The aseptic transfer of a part of a stock culture to a fresh growth medium.
2. A culture derived from a stock culture using the subculturing technique.

system disk

A *disk* which is needed in the start-up procedures in a computer.

system manager

A person responsible for the supervision and management of a *documentation system.* The person responds to comments and suggestions from users of the system on its efficiency and ease of use, and deals with any problems that arise.

T

table
A set of *data (e.g.* a file), in which each item *(e.g.* a *record)* is uniquely identified by its relative position in the set or by a label.

tape streamer
A device which uses magnetic tape for backing up *hard disks.* Commonly used for archival storage of large *data files* and as a safeguard against potential loss.

trait
A recognisable quality or attribute resulting from the interaction of a *gene* or group of genes with the environment.

U

uninterruptable power supply
Acts as an *interface* between electrical equipment and the main electricity source. When the uninterruptable power supply (or UPS) detects a drop in power or a loss of power altogether, it immediately steps in and supplies power from its own batteries.

UPS
See *uninterruptable power supply*

V

variation
Differences among individuals, populations or *species* in the expression of *traits* caused by both genetic and environmental factors.

VDU
Video display unit. Synonymous with *monitor.*

VGA
Video graphics array. See *monitor.*

viability
1. The capacity of an organism to live or to continue developing.
2. In seeds: The capacity to germinate once dormancy is removed.

video graphics array
See *monitor.*

view
The *fields* available to the user when a file or linked files are used. See also *screen form* and *mask.*

virus
A *program* which spreads between computers, usually via *floppy disks* or over *networks,* and often causes damage to computer *files.* Viruses can sometimes have devastating effects, destroying files and disrupting the operation of the computer.

video display unit
Synonymous with *monitor.*

voltage regulator
Designed to protect electrical equipment by guarding against any fluctuations in the power supply such as a surge or drop in power. See also *uninterruptable power supply.*

W

working collection
A *collection* of *germplasm* kept under short-term storage conditions. They are commonly used by breeders or researchers. Long-term maintenance is not a priority in working collections.

WORM
Write once, read many. A specialised compact *disk* which allows *data* to be written to the disk once and read for an unlimited number of times thereafter. See also *CD-ROM.*

write-protect
Any way of making files or disks *read-only.*

Bibliography

General genetic resources references

Breese, E.L. 1989. *Regeneration and multiplication of germplasm resources in seed genebanks: the scientific background.* IBPGR, Rome. 69 pp.

Brown, A.H.D., Clegg, M.T., Kahler, A.L. and Weir, B.S. (eds.). 1989. *Plant population genetics, breeding, and genetic resources.* Sinauer Associates, Sunderland, Massachusetts. 449 pp.

Brown, A.H.D., Frankel, O.H., Marshall, D.R. and Williams, J.T. 1989. *The USE of plant genetic resources*. Cambridge University Press, Cambridge, New York, New Rochelle, Melbourne, Sydney. 382 pp.

Douglas, J.E. 1980. *Successful seed program: a planning and management guide.* Westview Press, Boulder, Colorado.

Frankel, O.H. and Bennett, E. (eds.). 1970. *Genetic resources in plants – their exploration and conservation.* Blackwell Scientific Publications, Oxford and Edinburgh. 554 pp.

Frankel, O.H. and Hawkes, J.G. (eds.). 1975. *Crop genetic resources for today and tomorrow.* Cambridge University Press. Cambridge, London, New York, Melbourne. 492 pp.

Holden, J.H.W. and Williams, J.T. (eds.). 1984. *Crop genetic resources: conservation and evaluation.* George Allen and Unwin, London. 296 pp.

Hoyt, E. 1992. *Conserving the wild relatives of crops. [La conservation des plantes sauvages apparentées aux plantes cultives.] [Conservando los parientes silvestres de las plantas cultivadas.]* IBPGR, Rome/IUCN, Gland/WWF, Gland. 52 pp.

International Board for Plant Genetic Resources. 1990. *Elsevier's dictionary of plant genetic resources*. Elsevier Science Publishers, Amsterdam. 187 pp.

Keystone Center. 1991. Keystone international dialogue series on plant genetic resources. Oslo plenary session. *Final consensus report: global initiative for the security and sustainabIe use of plant genetic resources.* Genetic Resources Communications Systems, Washington. 42 pp.

National Research Council. 1991. *Managing global genetic resources – forest trees.* Committee on managing global genetic resources: Agricultural imperatives. National Academy Press, Washington, D.C. 228 pp.

National Research Council. 1991. *Managing global genetic resources – the U.S. national plant germplasm system.* Committee on managing global genetic resources: Agricultural imperatives. National Academy Press, Washington, D.C. 415 pp.

Paroda, R.S. and Arora, R.K. 1991. *Plant genetic resources: conservation and management.* Concepts and approaches. IBPGR, New Delhi. 392 pp.

Plucknett, D.L., Smith, N.J.H., Williams, J.T. and Anishetty, N.M. l987. *Gene banks and the world's food*. Princeton University Press, Princeton, New Jersey. 247 pp.

Withers, L.A. and Williams, J.T. 1982. *Crop genetic resources: the conservation of difficult material.* Proceedings of an international workshop help at the University of Reading, UK, 8-11 September 1980. I.U.B.S. Série B 42.

Witt, S.C. 1985. *Biotechnology and genetic diversity.* California Agricultural Lands Project.

Seed handling procedures

Ellis, R.H., Hong, T.D. and Roberts, E.H. 1985. Handbooks for Genebanks No. 2. *Handbook of seed technologies for genebanks. Vol. I. Principles and methodology.* IBPGR, Rome. 210 pp.

Ellis, R.H., Hong, T.D. and Roberts, E.H. 1985. *Handbooks for Genebanks No. 3. Handbook of seed technologies for genebanks. Vol. II. Compendium of specific germination information and test recommedations*. IBPGR, Rome. 456 pp.

Hanson, J. 1985. *Procedures for handling seeds in genebanks*. Practical Manuals for Genebanks (IBPGR). No. 1. IBPGR, Rome. 115 pp.

Characterisation and evaluation

Chapman, C. 1989. Principles of germplasm evaluation. In Stalker, H.T. and Chapman, C. (eds.). *Scientific management of germplasm: characterization, evaluation and enhancement.* IBPGR Training Courses: Lecture Series (IBPGR). No. 2. IBPGR, Rome, pp. 55-63.

Damania, A.B. 1985. Preliminary evaluation of *Triticum aestivum L.* from Nepal. *FAO/IBPGR Plant Genetic Resources Newsletter* 61:19-22.

Draper, S.R. and Keefe, P.D. 1989. Machine vision for the characterization and identification of cultivars. *Plant Var. Seeds 2:* 53-62.

Erskine, W. and Williams, J.T. 1980. The principles, problems and responsibilities of the preliminary evaluation of genetic resources samples of seed-propagated crops. *FAO/IBPGR Pl. Genet. Resources Newsl. 41:* 19-33.

Gerloff, J.E. and Smith, OS. 1988. *Choice of method for identifying germplasm with superior alleles.* 2. Computer simulation results. *Theor. Appl. Genet. 76:* 217-227.

Gupta, A.K., Saini, R.G., Shukla, K.K., Seth, D. and Purnima. 1981. Cataloguing wheat germplasm for disease resistance (in the Punjab Agricultural University, Ludhiana, India). *FAO/IBPGR Pl. Genet. Resources Newsl. 45:* 8-10.

Hayward, M.D. 1987. Isoenzymes as genetic markers for characterization of variability. *In* Tyler, B.F. (ed.). *Collection, characterization and utilization of genetic resources of temperate forage grass and clover.* IBPGR Training Courses: Lecture Series (IBPGR). No. 1. IBPGR, Rome, pp. 54-57.

Keefe, P.D. and Draper, S.R. 1988. An automated machine vision system for the morphometry of new cultivars and plant genebank accessions. *Plant Var.* Seeds *1:* 1-11.

Mloza-Banda, H.R. and Ayeh, E. 1988. Collecting, characterizing and conserving Malawi *Phaseolus vulgaris* germplasm. *FAO/IBPGR Pl. Genet. Resources Newsl. 75-76:* 39.

Pollock, C.J., Stoddart, J.L., Thomas, H. and Jones, T.W.A. 1987. Protein electrophoresis: isoenzymes as genetic markers. *In* Tyler, B.F. (ed.). *Collection, characterization and utilization of genetic resources of temperate forage grass and clover.* IBPGR Training Courses: Lecture Series (IBPGR). No. 1. IBPGR, Rome, pp. 41-53.

Rhodes, I. 1987. Characterisation of white clover. In Tyler, B.F. (ed.). *Collection, characterization and utilization of genetic resources of temperate forage grass and clover.* IBPGR Training Courses: Lecture Series (IBPGR). No. 1. IBPGR, Rome, pp. 18-24.

Simpson, M.J.A. and Withers, L.A. 1986. *Characterization of plant genetic resources using isozyme electrophoresis: a guide to the literature.* IBPGR, Rome, 102 pp.

Smith, J.S.C. 1989. Gene markers and their uses in the conservation, evaluation, and utilization of genetic resources of maize *(Zea mays* L.). *In* Stalker, H.T. and Chapman, C. (eds.). *Scientific management of germplasm: characterization, evaluation and enhancement.* IBPGR, Rome, pp. 125-135.

Thomas, H. and Gay, A.P. 1987. Characterization of forages for drought resistance. *In* Tyler, B.F. (ed.). *Collection, characterization and utilization of genetic resources of temperate forage grass and clover.* IBPGR Training Courses: Lecture Series (IBPGR). No. 1. IBPGR, Rome, pp. 29-35.

Yndgaard, F. and Hoskuldsson, A. 1985. Electrophoresis: a tool for genebanks. *FAO/IBPGR Pl. Genet. Resources Newsl. 63:* 34-40.

Genetic resources documentation systems – general

Alvarez, J., Levins, R.A. and Smiley, S.M. 1985. *Microcomputers as management tools in the sugar cane industry.* Elsevier Science Publishers, Amsterdam. 205 pp.

Blixt, S. 1984. Application of computers to genebanks and breeding programmes. *In* Vose, P.B. and Blixt, S.G. (eds.). *Crop breeding, a contemporary basis.* 1st ed. Pergamon Press, Oxford, pp. 293-310.

Blixt, S. 1988. Computer-supported gene bank management. *Kulturpflanze 36:* 121-134.

Bare, I. and Korpík, M. 1974. The results of a study on computer for the registration of genetic resources of the world collection of cultural plants in Czechoslovakia. Ustred-Vyzk-Ustav-Rostl-Vyroby, *-Ved-Pr 18:* 111-118.

Chardon, C. 1989. More efficient software to aid cane breeding program. *Bur. Sugar Exp. Sta. Bull. 27:* 8-9.

Hanson, J. and Watkins, R. 1978. Genetic resources information and the computer. *FAO/IBPGR Pl. Genet. Resources Newsl. 34:* 21-24.

Konopka, J. 1982. A computerized documentation system (for genebanks). In Blixt, S. and Williams, J.T. (eds.). *Documentation of Genetic resources: a model.* Workshop on Documentation Modelling for *Pisum* Genetic Resources. Alnarp (Sweden). 23 March 1982. IBPGR, Rome, pp. 79-81.

Konopka, J. and Hanson, J. 1985. Information handling systems for genebank management. *In* Konopka, J. and Hanson, J. (eds.). *Proceedings of a Workshop held at the Nordic Gene Bank,* Alnarp, Sweden, 21-23 November 1984. IBPGR, Rome, pp. 21-28.

Konopka, J. and Serwinski, J.T.I. 1982. Bazydanych W systemach dokumentacji zasobow genowych. [Data bases in documentation systems for genetic resources.] *Biul. Inst. Hodowli Aklim. Rosl. 147:* 159-164.

Mowder, J.D. and Stoner, A.K. 1989. Information systems. *Plant Breed.* Rev. *7:* 57-65.

Mowder, J.D. and Stoner, A.K. 1989. Plant germplasm information systems. *In* Knutson, L. and Stoner, A.K. (eds.). *Biotic diversity and germplasm preservation, global imperatives.* Kluwer Academic Publishers, Boston, pp. 419-426.

Witcombe, J.R. and Erskine, W. 1984. Documentation of germplasm collections by computer. *Adv. Agric. Biotechnol. 6:* 39-50.

Zhang, X.Z., Kong, F.S. and Yang, M.K. 1990. A design scheme for a database system of crop germplasm resources. *Scientia Agric. Sinica. 23:* 79-87.

Genetic resources documentation systems – examples

Andrews, D.J. and Hardwick, R.C. 1982. A database management system for information retrieval and documentation of experiments for plant breeders. *Euphytica 37:* 281-285.

Baum, B.R. 1991. Computerized world registries for cultivars. *Plant Breed. Abstr. 61:* 505-508.

Bencat', F. 1988. Koncepcia na vytvorenie databank genofondu rastlin na uzemi Slovenska. [A proposal for creating a databank of the plant gene pool in Slovakia..] *Biologia (Bratislava) 43:* 481-485.

Blixt, S. 1982. The pea model for documentation of genetic resources. *In* Blixt, S. and Williams, J.T. (eds.). *Documentation of genetic resources: a model.* Workshop on Documentation Modelling for *Pisum* Genetic Resources. Alnarp (Sweden). 23 Mar 1982. IBPGR, Rome, pp. 3-28.

Burillon, G. 1991. La base de données du Laboratoire de Graines du CTFT. FAO. *lnform. Ressources Genet. For. 18:* 45-47.

Commonwealth Scientific and Industrial Research Organization, Canberra. Div. of Forest Research. 1991, Treeseed – a management system for the maintenance of tree seed collections. FAO. *For. Genet. Resources Inf. 18:* 48-49.

Elgin, J.H. Jr, Alexanian, S.M and Aleksanyan, S.M. l991. Plant germplasm accession, maintenance and data management at the Vavilov Institute of Plant Industry (VIR), Leningrad, USSR. *Plant Breed. Abstr. 61:* 397-402.

End, M.J., Wadsworth, R.M., Hadley, P. 1990. The primitive cocoa germplasm database. *Cocoa Growers' Bull.* 1990, *43:* 25-33.

Engels, J.M. 1985. Documentation and information management at GCRC/E. PGRC/E ILCA *Germplasm Newsl. 9:* 20-27.

Food and Agriculture Organization of the United Nations: 199l. El estado de los recursos fitogenéticos en el mundo y el Sistema Mundial de Información y Alerta sobre Recursos Fitogenéticos. Dirección de Producción y Protección Vegetal. Comisión de Recursos Fitogenéticos. Reunión. 4. Rome. 15-19 Apr 1991. FAO, Rome. 10 pp.

Fogle, H.W. and Winters, H.F. 1974. Computerized fruit germplasm resources inventory. *Fruit Var. J. 28:* 35-37.

Frese, L. and van Hintum, T.J.L. 1989. *The International Data Base for* Beta. International Conference on *Beta* Genetic Resources. Wageningen (Netherlands). 7-10 February 1989. *In* International Crop Network Series (IBPGR). No. 3. IBPGR, Rome, pp. 17-45.

Gomez, K.A., Chang, T.T., Francisco, C.J. and Chan, M .A. 1978. *The computerized management system for the IRRI Germplasm Bank.* Proceedings of Workshop on the Genetic Conservation of Rice. IRRI/IBPGR, Los Baños, Philippines. 26 pp.

Goonewardene, H.F. 1986. *A computer management system for apple ("Malus X domestica" Borkh.)*

germplasm with resistance to disease and arthropod pests. USDA, Springfield, Virginia. 26 pp.

Gullberg, U. 1973. Computerization of grass information in Kenya. In *Proceedings of a Conference on European and Regional gene banks.* European Association for Research on Plant Breeding, Izmir, pp. 57-58.

Hazekamp, Th. and van Hintum, Th.J.L. 1990. GENIS-PCC: The documentation of dispersed collections. *FAO/IBPGR Pl. Genet. Resources Newsl. 81/82:* 13-15.

International Rice Research Institute. 1976. Computerized data management. In *IRRI Annual Report for 1976.* IRRI, Los Baños, Laguna, Philippines, pp. 134-135.

Iyama, S. c.1988. Management of genetic resources information in Japan. In Suzuki, S. (ed.). *Crop genetic resources of East Asia.* Proceedings of the International Workshop on Crop Genetic Resources of East Asia. Tsukuba, Japan, 10-13 November 1987. IBPGR, Rome, pp. 247-248.

Kawakatsu, M. l987. Utilization of a database on sugarbeet breeding. 2. Retrieving items from a database constructed by using GRIMS/CGS. *Proc. Sugar Reef Res. Assoc., Japan 29:* 1-8.

Kawakatsu, M. and Amma, S. l986. Utilization of a database on sugarbeet breeding. 1. Construction of a database using GRIMS/CGS. *Proc. Sugar Reef Res. Assoc. Japan 28:* 44-50.

Kirchheim, B. and Kranz, A.R. l985. Computerized listing of the AIS-seed bank material. *Arabidopsis Inf. Serv. 22:* 147-152.

Knüpffer, H. 1988. The European Barley Data Base of the European Cooperative programme for the Conservation and Exchange of Crop Genetic Resources (ECP/GR). *FAO/IBPGR Pl. Genet. Resources Newsl. 75/76:* 17-20.

Lu, C.H., Huang, Z.M., Cao, J.S., Jiang, H., Wu, J.L. and Wang, G.L. 1987. Establishment of the database for rice germplasm in Taihu Lake area and its application. *In* Proceedings of the 1st International Conference on Agricultural Systems Engineering, 11-14 August l987, Changchun, China. pp. 567-577.

Martynov, S.P. 1991. Computer bank of passport data on wheat. *Selek. Semenovod. (Moscow) 1:* 14-16.

McMillan, C. and Salhuana, W. c.1983. lnformation management systems for forage plant genetic resources. In McIvor, J.G. and Bray, R.A. (eds.). *Genetic resources of forage plants.* CSIRO, Melbourne, pp. 299-308.

Pedersen, A.P. 1991. Description generale de "Seed Bank", système de base de données developpé et utilisé par le Centre de Semences Forestières DANIDA (Denmark). *FAO. lnformations sur les Ressources Genetiques Forestieres. 18:* 42-44.

Perry, M.C. 1988. The Germplasm Resources Information Network – A genetic resources information management system for the United States. *FAO/IBPGR Pl. Genet. Resources Newsl. 73/74:* 21-26.

Porceddu, E. 1973. Handling of information on genetic resources at the Italian germplasm station. In *Proceedings of a Conference on European and Regional gene banks.* European Association for Research on Plant Breeding, Izmir, pp. 48-50.

Porter, W.M. and Smith, D.H., Jr. 1982. Computer-assisted management of the USDA small grain collection. *Plant Dis. 66:* 435-438.

Riemenschneider, D.E. 1983. Alias – a computer system for tracking breeding materials with several names. *In* Proceedings bf the North Central Tree Improvement Coliference, Madison, Dept. of Forestry, University of Wisconsin, Madison, pp. 123-127.

Rogalewicz, V. and Bares, I. 1987. Czechoslovak information system for genetic resources. *FAO/IBPGR Pl. Genet. Resources Newsl. 72:* 14-16.

Rogers, D.J. l973. The concepts and applicability of TAXIR for storage and retrieval of gene bank data. *In* Proceedings of a Conference on European and Regional gene banks. European Association for Research on Plant Breeding, Izmir, pp. 37-44.

Snoad, B. 1973. A computer-based recording system for *Pisum* accession and genetic data. *In* Proceedings of a Conference on European and Regional gene banks. European Association for Research on Plant Breeding, Izmir. pp. 55-56.

Suzuki, S. and Kumagai, K. 1981. Present status of genetic resource information management in Japan. *In* Proceedings of the Symposium on Technical Aspects of Plant Genetic Resources, Tsukuba University, Tsukuba, Japan, 25-26 October 1980, pp. l-17.

Thompson, B.K. and Baum, B.R. 1978. Preparation of a barley register: an information retrieval system for cultivars. *Taxon. 27:* 471-477.

Valmayor, R.V. 1977. Major activities in exploration, conservation, documentation and evaluation of plant genetic resources in the Philippines. FAO, Rome, IBPGR, Rome and Philippine Council for Agriculture and Resources Research, College, Laguna. *Report on Southeast Asian Workshop on Plant Genetic Resources.* Philippines, 6 December 1976, pp. 29-36.

van Hintum, T.J.L. 1988. GENIS: a fourth generation information system for the database management of genebanks. *FAO/IBPGR Pl. Genet. Resources Newsl. 75/76:* 13-15.

Vanderborght, T. ca.1988. A centralized database for the common bean and its use in diversity analysis. *In* Gepts, P. (ed.). *Genetic resources of Phaseolus beans: their maintenance, domestication, evolution, and utilization.* Kluwer Academic Publishers, Dordrecht, pp. 51-65.

Wheelans, K. and Withers, L.A. 1984. The IBPGR (International Board for Plant Genetic Resources) international database on *in vitro* COnservation. *FAO/IBPGR Pl. Genet. Resources Newsl. 60:* 33-38.

Wheelans, S.K. and Withers, L.A. 1988. The IBPGR *in vitro* conservation data bases. *NATO ASI Ser H Cell Biol. 18:* 497-500.

Yndgaard, F. 1982. A documentation system for the Nordic Gene Bank. *FAO/IBPGR Pl. Genet. Resources Newsl. 49:* 34-36.

Yndgaard, F. 1990. Computerized pedigree information on Nordic barley cultivars. *Sveriges Utsadesforen. Tidskr. 2:* 105-111.

Documentation procedures and concepts

Burley, J. and von Carlowitz, P. 1984. *Multipurpose tree germplasm.* Proceedings, Recommendations and Documents of a planning Workshop to discuss International Cooperation National Academy of Sciences, Washington, DC, USA, June 1983. International Council for Research in Agroforestry, Nairobi. 298 pp.

Chang, T.T. 1985. Evaluation and documentation of crop germplasm. *Iowa State J. Res. 59:* 379-397.

Creech, J.L. 1981. Practical questions regarding use and exchange of genetic resources and information. Crop genetic resources of the Far East and the Pacific. *In* Williams, J.T., (ed.). *Symposium on the Genetic Resources of the Far East and the Pacific Islands.* Tsukuba, Japan, 21-24 October 1980, IBPGR, Rome. pp. 153-155.

Esquinas-Alcazar, J.T. and Contreras, M.A. 1987. *Documentacion de los recursos fitogenéticos. Anales*. Simposio "Recursos Fitogenéticos". Valdivia (Chile). 20-22 November 1984. UACH, Valdivia/IBPGR, Rome, pp. 178-187.

Ford-Lloyd, B. and Jackson, M. 1986. The management of genetic resources data. *In* Plant genetic resources – an introduction to their conservation and use. Edward Arnold, London, pp. 69-80.

Hersh, G.N. and Rogers, D.J. 1975. Documentation and information requirements for genetic resources application. *In* Frankel, O.H. and Hawkes, J.G. (eds.). *Crop genetic resources for today and tomorrow,* International Biological Programme 2. Cambridge University Press, Cambridge, pp. 407-443.

Howes, C. 1981. Guidelines for developing descriptor lists. *FAO/IBPGR Pl. Genet. Resources Newsl. 45:* 26-32.

International Board for Plant Genetic Resources. 1984. *Exchange of information.* Report of a Workshop held at the plant Breeding and Acclimatisation Institute, Radzikow, Poland. 23-25 October 1984. UNDP-IBPGR, Rome, pp. 15.

Kolasinski, P. 1988. Notes on some descriptor states. FAO/IBPGR Pl. Genet. Resources Newsl. *75/76:* 25-36.

Konopka, J. 1985. Workshop on Exchange of lnformation. *FAO/IBPGR Pl. Genet. Resources* Newsl. *61:* 38.

Konzak, C.F. 1973. Standardized documentation procedures for germplasm collections. In *Proceedings of a Conference on European a and Regional Gene Banks.* European Association for Research on Plant Breeding, Izmir, pp. 28-36.

Konzak, CF., Walden, W.E. and Sousa, F.P. 1970. Problems and progress in the management of information on genetic resources. *Seiken Zihô 22:* 91-97.

Laboratory for Information Science in Agriculture. 1990. International conference acts to establish global center for plant germplasm information. *Diversity 6:* 5-6.

Lockerman, R.H., Bisby, F.A. and van der Maesen, L.J.G. 1988. Workshop: integration of information on plant diversity. *In* Summerfield, R.J. (ed.). *World crops: cool season food legumes: a global perspective of the problem and prospects for crop improvement in pea, lentil, faba bean and chickpea.* Proceedings of the International Food Legume Research Conference on Pea, Lentil, Faba Bean and Chickpea, held at the Sheraton Hotel, Spokane, Washington, USA 6-11 July 1986. Kluwer Academic Publishers, Boston, pp. 129-133.

Nguyen-Van, E. and Pernes, J. 1984. Les bases de données et leur exploitation statistique. In Pernes, J. (avec la collaboration de J. Berthaud et al.). *Gestion de ressourccs généfiques des plantes.* Agence de coopération culturelle et technique, Paris, pp. 235-291.

O'Brien, S.J. (ed.). 1990. *Genetic maps – Locus maps of complex genomes.* 5th edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, *New* York. 1103 pp.

Parliman, B.J., Talbot, H., Mowder, J.D. and Foran, N.F. 1985. Quarantined and quarantine-released *Prunus* sp. accessions: the integration of quarantined-germplasm inventories into the germplasm resources information network (GRIN). *Fruit Var. J. 39:* 7-11.

Peeters, J.P. and Williams, J.T. 1984. Towards better use of genebanks with special reference to information. *FAO/IBPGR PI. Genet. Resources Newsl. 60:* 22-32.

Powell, R.L. and Norman, H.D. 1989. Animal germplasm information systems. In Knutson, L. and Stoner, A.K. (eds.). *Biotic diversity and germplasm preservation, global imperatives.* Kluwer Academic Publishers, Boston, pp. 427-443.

Roelofsen, H. 1985. Using evaluation data: an information problem. *In* Rogalewicz, V. (ed.). *Proceedings of the EUCARPIA Genetic Resources Section International Symposium: Evaluation for the Better Use of Genetic Resources Materials* held at the Research Institute of Plant Production, Prague-Ruzyne, Czechoslovakia, March 27-29, 1985. Research Institute of Plant Production, Prague, pp. 167-173.

Roelofsen, H.J. 1986. Computer-aided classification of non-numerical gene bank data. Acta *Hortic. 182:* 309-317.

Rogalewicz, V. 1987. Possibilities and limits of germplasm information systems. *Sri. Agriculturae Bohemoslavaca 19:* 295-299.

Rogers, D.J., Snoad, B. and Seidewitz, L. 1975. Documentation for genetic resources centers. *In* Frankel, O.H. and.Hawkes, J.G. (eds.). *Crop genetic resources for today and tomorrow,* International Biological Programme 2. Cambridge University Press, Cambridge, pp. 399-405.

Seidewitz, L. 1976. *Thesaurus for the international standardisation of genebank documentation. Part VA: Selection of common scientific terms for plant pests and diseases.* Forschungstalt fuer Landwirtschaft Braunschweig-Völkenrode (FAL), Braunschweig.

Terrell, E.E. 1986. Updating scientific names for introduced germplasm of economically important vascular plants. *Acta Hortic. 182:* 293-300.

United States General Accounting Office. 1990. Report to the Secretary of Agriculture. *Plant Germplasm – Improving data for management decisions. Vol. 1.* GAO/PEMD-9l-5A. Washington, D.C. 69 pp.

United States General Accounting Office. 1990. Report to the Secretary of Agriculture. *Plant Germplasm – A data collection framework and questionnaire.* Vol. 2. GAO/PEMD-9l-5B. Washington, D.C. 83 pp.

van Hintum, T.J.L. 1989. De beste genotypeschatters bij zeer incomplete waarnemingensets. [The best genotype evaluators for use with very incomplete sets of observations.] *Prophyta. 43:* 37-39.

van Hintum, T.J.L. 1989. *Strategies for selecting subsets within collections.* International Conference on *Beta* Genetic Resources. Wageningen (Netherlands). 7-10 February 1989. In International Crop Network Series (IBPGR). No. 3. IBPGR, Rome, pp. 82-88.

Vaughan, D.A. 1991. Choosing rice germplasm for evaluation. *Euphytica 54:* 147-154.

van Loosdrecht, M.P.H., von Soest, L.J.M. and Dik, M.T.A. 1988. Descriptor list for wheat developed for a regeneration programme in the Netherlands. *FAO/IBPGR Pl. Genet. Resources Newsl. 73/74:* 35-38.

Watson, L., Aiken, S.G., Dallwitz, M.J., Lefkovitch, L.P. and Dubé, M. 1986. Canadian grass genera: keys and descriptions in English and French from an automated data bank. *Can. J. Bot. 64:* 53-70.

Williams, J.T. 1982. Information needs in crop genetic resources work. *In* Blixt, S. and Williams, J.T. (eds.). *Documentation of genetic resources: a model.* Workshop on Documentation Modelling for *Pisum* Genetic Resources. Alnarp (Sweden). 23 March 1982. IBPGR, Rome, pp. 29-31.

Williams, J.T. 1986. Classification of genetic resources samples and breeders' stocks. In Styles, B. (ed.). *Infraspecific classification of wild and cultivated plants.* Clarendon Press, Oxford, pp. 225-261.

Descriptor lists published by IBPGR

Almond (Revised). 1985.
Annual *Medicago/Medicago* annuelles. 1991
Apricot. 1984.
Apple. 1982.
Bambara groundnut. 1987.
Banana (revised). 1984.
Barley. 1982.
Beta. 1991.
Brassica and *Raphanus.* 1990.
Brassira campestris L. 1987.
Cashew. 1986.
Cherry. 1985.
Chickpea. 1985.
Citrus. 1988.
Colocasia. 1980.
Cotton (revised). 1985.
Cowpea. 1983.
Cultivated and wild sunflower. 1985.
Echinochloa millet. 1983.
Eggplant/Aubergine. 1990.
Faba bean. 1985.
Finger millet. 1985.
Forage grass. 1985.
Forage legumes. 1984.
Grape. 1983.
Groundnut (revised). 1985.
Kodo millet. 1983.
Lentil. 1985.
Lima bean. 1982.
Lupin/Lupinos. 1981.
Maize/maiz/mais. 1991.
Mango. 1989.
Mung bean. 1980
Oat. 1985.
Oca. 1982.
Oilpalm. 1989.
Panicum miliaceum and *P. sumatrense.* 1985.
Papaya. 1988.
Peach. 1985.
Pear. 1983.
Pearl millet. 1981.
Phaseolus acutifolius. 1985.
Phaseolus coccineus. 1983.
Phaseolus vulgaris. 1982.
Pineapple. 1991.
Pigeonpea. 1981.
Plum. 1985.
Potato (cultivated). 1977.
Quinoa. 1981.
Rice. 1980.
Rye and *Triticale.* 1985.
Safflower. 1983.
Sesame. 1981.
Setaria italica and *S. pumila.* 1985.
Sweet potato/batata/patate douce. 1991.
Sorghum (revised). 1984.
Soyabean. 1984.
Strawberry. 1986.
Sunflower. 1985.
Tropical fruits. 1980.
Vigna aconitifolia and *V. trilobata.* 1985.
Vigna mungo and *V. radiata* (revised). 1985.
Wheat and *Aegilops.* 1989.
Winged bean (revised). 1982.
Xanthosoma. 1989.
Yam *(Dioscorea* sp.). 1980.

Directories of germplasm collections published by IBPGR

Bettencourt, E. and Konopka, J. 1988. Directory of germplasm collections. 5.II. *Industrial crops Beet, coffee, oil palm, cotton and rubber.* IBPGR. Rome. 61 pp.

Bettencourt, E., Konopka, J. and Damania, A.B. 1989. Directory of germplasm collections. 1.1. *Food legumes.* Arachis, Cajanus, Cicer, Lens, Lupinus, Phaseolus, Pisum, Psophocarpus, Vicia *and* Vigna. IBPGR, Rome. 190 pp.

Bettencourt, E. and Konopka, J. 1989. Directory of germplasm collections. *6.II. Temperate fruits and nuts.* Actinidia, Amerlanchier, Carya, Castanea, Corylus, Cydonia, Diospyros, Fragaria, Juglans, Malus, Mespilus, Morus, Olea, Pistacia, Prunus, Pyrus, Ribes, Rosa, Rubus, Sambucus, Sorbus, Vaccinium *and others.* IBPGR, Rome. 296 pp.

Bettencourt, E. and Konopka, J. 1989. Directory of germplasm collections. *4. Vegetables.* Abelmoschus, Allium, Amaranthus, *Brassicaceae,* Capsicum, *Cucurbitaceae,* Lycopersicon, Solanum *and other vegetables.* IBPGR, Rome. 250 pp.

Bettencourt, E. and Konopka, J. 1990. Directory of germplasm collections. *3. Cereals.* Avena, Hordeum, *Millets,* Oryza, Secale, Sorghum, Triticum, Zea *and pseudocereals.* IBPGR, Rome. 264 pp.

Davies, W.E. and McLean, B.T. 1984. *Directory of germplasm resources: forages (grasses, legumes, etc).* IBPGR, Rome. 47 pp.

Gulick, P. and van Sloten, D.H. 1984. Directory of germplasm collections. *6. I. Tropical and sub-tropical fruits and free nuts.* IBPGR, Rome. 191 pp.

Juvik, G.A., Bernard, R.L. and Kauffman, H.E. 1985. Directory of germplasm collections: 1.II. *Food legumes (soyabean).* IBPGR, Rome. 59 pp.

Lawerence, T., Toll, J. and van Sloten, D.H. 1986. Directory of germplasm collections. 2. *Root and tuber crops.* IBPGR, Rome. 178 pp.

Williams, J.T. and Damania, A.B. 1981. Directory of germplasm collection. 5. *Industrial crops. I. Cacao, coconut, pepper, sugarcane and tea.* IBPGR, Rome. 54 pp.

Answers to exercises

ANSWERS

Chapter 2

1. a . true
   b . false
   c . false
   d . true
   e . false
   f . true
   g . false
   h . true – but can also maintain fruit trees and glass house crops
   i . true

2. Genebanks are genetic resources centres which are set up principally to conserve germplasm in one or more collections. The germplasm conserved can be seeds, tissue culture or actively growing plants.

3. Genebanks differ from one another in their activities and how the activities are organised and performed with the resources available. These activities are themselves determined by the overall aim of the genebank and the objectives in the different subject areas.

4. A discussion about the different collections can be found in section 1.5.

5. a . true
   b . false
   c . false
   d . false
   e . true
   f . false
   g . true
   h . false
   i . true

6. An explanation of the difference between data and information can be found in section 2.1.

ANSWERS

7. Priorities need to be set to maximise the use of the limited resources and the limited time available.

8. A discussion about the role of information in genebank management can be found in section 2.2.

9. A genebank documentation system is tailor-made according to the documentation and information needs of the genebank. Since genebanks differ from one another according to their activities and how they are organised, it follows that their documentation systems will also be different.

10. A description of desirable features can be found in section 2.4.

11. Related data are organised into groups which are practical to use in terms of data recording, storage and maintenance. These groups are closely associated with the genebank procedures.

12. A detailed discussion about the stages in construction is given in section 2.5

Chapter 3

1. a. false
   b. true
   c. false
   d. false
   e. true
   f. false
   g. true
   h. false
   i. true

2. Accession specific data are concerned with individual accessions whereas group data are concerned with groups of accessions. Examples of the former include seed weight, moisture content and flower colour; examples of the latter include equilibrium moisture contents, literature references and viability test methods.

3. Accession specific data is given a higher priority for documentation than group data as these data are important for genebank management

4. A detailed discussion of the different approaches to accession numbering systems is given in section 1.1. In most cases it is best to operate a single, numeric accession numbering system in the genebank which is strictly sequential in its operation.

5. The batch reference is routinely recorded since many genebank operations are performed on germplasm from specific regeneration cycles. In characterisation and preliminary evaluation there is a possibility of variation between different regeneration cycles, and so the batch reference should be recorded. Also, in seed store inventories, each batch can have different viabilities, moisture contents, amount of seed store and so on. As it is important to keep track of these data for management purposes, the batch reference is routinely recorded.

6. a . true
   b . false
   c . true
   d . false
   e . true
   f . false
   g . true
   h . true
   i . false

7. A well designed procedure makes a task easier to carry out, easier to manage and makes the best use of available resources. Fewer mistakes are made and any data produced are more reliable.

8. An operational procedure is one concerned with the day to day running of the genebank and produces data of a high management value. Commonly, these data need to be updated at a later date. Examples include seed drying, seed viability testing and seed storage. A scientific procedure is one which produces data of potential interest to those outside of the genebank such as characterisation or preliminary evaluation trials.

9. Data from operational procedures tend to have a high management value. For this reason, a high priority is given for their documentation.

10. Flow charts are used get an overall picture of the different steps in a procedure, the decisions taken at each step, the data relationships and the relationship with other procedures. A discussion of the design considerations is given in section 4.

Chapter 5

1. a . yes
   b . no
   c . yes

ANSWERS

d . yes
e . yes
f . yes
g . no
h . no – see comments in 3 (i) below
i . yes
j . no
k . no

2. ordinal: b., c., f., g., h.
nominal: a., d., e., i., j.

3. (i) The following can be scored using both scales: a., c., e., f., g., i., j., k., l., m., n., o., p., q. and r.. Descriptors b., d., h., s. and t. are scored using qualitative scales. It should be pointed out that although the descriptor "Accession number" is often numerical, it is used as a label and therefore is qualitative; you cannot, say, use the accession number meaningfully in a calculation.

3. (ii) All quantitative data can theoretically be scored on an ordinal scale but as there is *always* a loss in precision this can often be unacceptable. It is recommended therefore that all raw quantitative data are scored on a continuous scale to maximise the information potential of the data. These data can be converted at a later date to an ordinal scale if desired.

4. Collected from a wild habitat
Collected from a farm land
Collected from a farm store
Collected from a backyard
Collected from a village market
Collected from a commercial market
Collected from an institute

It is better to score this descriptor using a nominal scale because each accession has only one possible collecting source. If a binary scale were used, 6 of the above descriptors would be redundant for each accession.

5. a. This scale should be rearranged as follows:

l=very short (<0.5m)
2=very short-short (0.5-0.75m)
3=short (>0.75-1.0m)
4=short to intermediate (>1.0-1.25m)
5=intermediate (>1.25-1.5m)
6=intermediate to tall (>1.5-1.75m)
7=tall (>1.75-2.0m)
8=tall to very tall (>2.0-2.25m)
9=very tall (>2.25m)

b. "Absent" should always be scored with the character zero "0". The scale should therefore be re-written as follows:

0=none
3=few
5=medium
7=abundant

c. When using a scale of 1-9 for ordinal scale data, the convention is that 1=very low and 9=very high. The scale should therefore be re-written as follows:

1=very loose (spongy, large air cavities)
3=loose (crumbly)
5=average
7=dense (fleshy)
9=very dense (very fleshy)

d. The character zero "0" should be reserved for "absent" or "not observed". The scale should therefore be re-written as follows:

1=yellow
2=orange
3=red
4=dark purple
5=purple

e. As it stands, this is not a binary scale at all but a nominal scale using the wrong convention for the character zero. It should be re-written as: 1=grouped fruitlets, 2=isolated fruitlets. To be a binary scale it needs to be split into two separate descriptors and the correct convention used for absent and present, namely:

Presence of grouped fruitlets
+=present
0=absent

Presence of isolated fruitlets
+=present
0=absent

Chapter 6

1. a. false
   b. true
   c. false
   d. true
   e. true
   f. false

ANSWERS

g. false
h. false
i. false

2. Occasions when it is not possible to record data directly include those when it is impractical (the computer is stored in a different location), where data analysis is required before formal documentation and when several people need access to the documentation system at the same time.

3. Clear layouts are the key to user-friendly forms; they should be easy to fill in, easy to update and easy to read. To this end pre-printed forms should be used with columns (or boxes) wide enough to accommodate the data and the columns thoughtfully arranged to facilitate data recording and retrieval. Where there are many descriptors, a landscape orientation should be used or more than one form designed to accommodate them. The form should have a title and there should be a separate column for any comments.

4. The stages are:
   1. Decide whether to use a management file
   2. Develop manual forms for recording raw data and scoring descriptors
   3. Organise the forms into separate files according to subject area
   4. Develop documentation procedures to ensure that the manual documentation system is operated correctly

5. A management file should contain those descriptors which are essential to the management of the collections (e.g. descriptors for viability, amount of germplasm, location in store; date of next viability test etc.) Discrepancies can be avoided if the descriptors are scored only in the management file and in no other place in the documentation system.

6. Data should be arranged in an order which facilitates data recording and information retrieval. If an unsuitable order were chosen, unnecessary work is created entering data, keeping the system up to date and retrieving specific data.

7. One way to deal with feedback information is to keep an information sheet for each accession which details all literature references, anecdotal remarks and specific information (e.g. from field trials).

8. Raw data should not be discarded as they can be useful for future analysis or for background information (e.g. to find out the precise conditions of an evaluation trial). To keep track of raw data, all diaries, notebooks and sheets should be stored according to the subject area and date. Also, the source of the raw data can be stored in the documentation system so that they can be accessed at a later date.

ANSWERS

Chapter 7

1. a. true
   b. true
   c. false
   d. false
   e. true
   f. true
   g. true
   h. false
   i. false
   j. false

2. a. memory
   b. CPU *or* central processing unit
   c. mouse
   d. hard disk
   e. CD-ROM's
   f. modem
   g. boot sector
   h. floppy disks

3. Computer hardware commonly comprises a monitor, keyboard, system unit, printer and a mouse. The keyboard is used to type in instructions or text into the computer. Alternatively, instructions can be entered with the mouse. The monitor displays the output from programs or what has been typed in. The printer is used for printing documents.

4. Hard disks have a much higher storage capacity than floppy disks and operate more quickly. Most hard disks are stored inside the system unit and, unlike floppy disks, are not removable. Other storage devices include removable cartridges (which have a similar capacity to hard disks), CD-ROMs (very large capacity) and tape streamers which are used for archival storage of large data files.

5. There are four types of printer commonly available: daisy wheel, dot matrix, ink jet and laser. Ink jet and laser printers are the most appropriate printers for high quality graphics; daisy wheel printers cannot print graphics and generally the quality of the output for dot matrix printers is not high-enough.

6. Fluctuations in power can cause erratic behaviour in the hardware leading damaged files, data corruption and loss of work. In extreme cases, serious and extensive damage to the hardware can be caused. To avoid this, a voltage regulator can be installed to protect the sys-

ANSWERS

tem from fluctuations in power and an uninterruptable power supply used to protect the system against power cuts.

7. Other ways of entering data include the use of pen-input devices (e.g. with a graphics tablet), scanners (for the input of text and graphics) and the use of modems for the input and output of data over a telephone line.

8. File viruses are viruses which infect program files and activate when the program is run. They can spread easily to other files. A boot sector virus infects the boot sector hard or floppy disk and are spread easily as they are carried by the operating system. Worms are viruses which infect computer networks, bombs are viruses activate on a particular date and trojan horses which are not viruses in the strict sense but seemingly innocent programs which cause severe damage when operated.

9. Possible reasons are as follows:
   1. Wrong size of disk e.g. 5¼" rather than 3½"
   2. Different storage capacity of disk (e.g. high storage capacity) which is not recognised by another computer
   3. Disk was formatted for a different operating system
   4. Programs on the disk were written for a different microprocessor/operating system combination.

10. A list of precautions you should take can be found in section 7.

ANSWERS

Chapter 8

1. a. false
   b. true
   c. false
   d. true
   e. true
   f. false
   g. false
   h. true
   i. false
   j. false

2. a. field
   b. identifying fields
   c. index
   d. character
   e. logical

ANSWERS

3. Flat file managers are simple database managers which work with one file at a time. In more sophisticated flat file managers, different files can make reference to each other and therefore worked with as a single unit. Relational database managers are designed to work with more than one file at a time and store the same data element in one place only.

4. Two examples:
   a. Descriptors for distribution of seed from the seed store are concerned with details about the accessions sent and the recipients of the germplasm; these should normally be stored in separate files which can be linked together.
   b. Important management data (e.g. inventory, seed viability and seed drying data) which can be placed in a single file in a manual documentation system should normally be stored in separate linked files in a computerised system.

5. A relationship is outlined in section 3.

6. A register of field descriptions (or data dictionary) should contain details on the descriptor name, the field name and description, the field type, field width and any index used. Recording these details will help ensure consistency between files and make the process of building the documentation system much easier.

7. Some of the descriptor names are very long and there are several possible abbreviations that you could use. The following are therefore only suggestions:

   a. acc_no
   b. acq_date
   c. altitude
   d. coll_inst
   e. coll_name
   f. coll_no
   g. coll_date
   h. no_plants
   i. local_name
   j. stem_pig
   k. eval_env
   1. loc_store
   m. frost_susc
   n. name_pers
   o. soil_ph
   p. poll_meth
   q. next_test
   r. plot_no

s. harvest_mc
t. storage_mc
u. lot_duplic

8. Spreadsheets, despite being powerful tools for performing calculations, lack the power of database management software for flexible data management. They can still be used in conjunction with database management software data are exported from a database to a spreadsheet for further analysis.

Chapter 9

1. a. false
   b. true
   c. false
   d. false
   e. true
   f. true
   g. false
   h. true
   i. false
   j. true

2. The physical system design is a detailed plan of how the documentation system will operate in practice with the chosen hardware/software combination and the users of the system. The logical system design is an outline of how the system should operate in theory based on user requirements without reference to hardware or software. The logical system design has to be modified if some of the procedures outlined are impractical (e.g. the software is not powerful enough, insufficient computers available etc.).

3. When designed in an appropriate way, screen forms can help prevent errors getting into the documentation system and make data entry and modification much more user-friendly. Screen forms should be easy to read – avoiding the use of abbreviations and with not too many descriptors per form. There should be consistency in design between different forms and, where appropriate, the order of the fields on the screen should be the same as for the corresponding manual form. It is also important to exploit the features of the software to trap errors at the point of entry and also to facilitate data entry.

4. A list of different features is given in section 2.2.6.

ANSWERS

5. *1.* The descriptors which appear in the report are identified
   2. The fields which are used in the search are identified
   3. The need for any links between files is established
   4. The search criteria are defined – in this case to select accessions with viabilities below a certain level
   5. The way in which the records should be displayed is defined – in this case sorted in order of crop and numerical order
   6. The layout of the report is designed; several designs are tried until a suitable one is found
   7. The final report is produced.

6. A discussion of these features is given in section 4.

7. The organisation of all the different routines into a. menu driven system considerably facilitates the use of the documentation system. Menus can be organised according to the data operation (enter, modify, search, report etc.) or the subject area (registration, passport, characterisation etc.).

8. The documentation system should have good system documentation to facilitate the use and future modification of the system. The different approaches are given in section 6.

Chapter 10

1. A system manager is a person responsible for the supervision and maintenance of a documentation system. The system manager should be able to deal with any problems in system, respond. to comments and suggestions from the users and handle requests for information.

2. A system can be introduced:
   a. As an immediate change-over – requiring extensive testing prior to the change-over to avoid mistakes
   b. In parallel with the old system – involving a large duplication of effort but avoiding some of the difficulties in the change-over period
   c. Over a period as separate modules – allowing flexibility and the ability to learn from mistakes throughout the process.

3. Training considerations are outlined in section 1.1

4. Software should always be purchased from a commercial supplier and, where possible, the disks should be write-protected before installing the software. Illegal copies of software should not be used. Anti-virus should be used to protect the system and to check any new disks which are used.

5. Data exchange is facilitated if the same database management software is used with the same file structure and data formats. The accuracy of data transfer should be thoroughly checked using test data. It is also important that back-up copies of files are made before data exchange takes place in case any problems arise.

Appendix I: Completed forms

This appendix provides completed examples of the questionnaire on pages 26-30 and the forms on pages 85 and 86 of this guidebook. You will have used the questionnaire to help you analyse your genebank as a whole, and the forms to help you analyse each of the individual procedures performed at your genebank. Try not to model your responses on those given in this appendix. The completed questionnaire and forms are provided only as a guide in case you are having difficulties answering the questions or performing the analysis of your procedures.

The organisation chosen as a case study for completing the questionnaire and the forms is the Genebank of Ornamental & Medicinal Plants (OMP). The OMP genebank belongs to the unit of Ornamental and Medicinal Plants, Department of Horticulture, in the Faculty of Agriculture of X University. It was established in 1964 to provide researchers of the university and associated breeders of ornamental and medicinal plants with stocks of primitive cultivars, mutants, lines and hybrids necessary for further research, seed and seedling production of these crops. The OMP genebank holds 1,200 accessions of various ornamental and medicinal plant species, mostly but not exclusively of endemic origin, such as *Rosa, Viola, Chrysanthemum, Oenothora, Mentha, Lawsonia, Geranium, Aloë* and *Thymus.*

The seed collection is maintained as "active collection" under storage conditions of 5°C with relative humidity control of 20-25% R.H. Sources of this collection include horticultural collecting missions of this university and other outside collectors, introductions from other research institutions abroad, breeders collections from other institutes, etc. Part of this collection is maintained in the field/orchard/glasshouse facilities. The staffing organisation of this genebank includes a curator (senior technician at the department) and six technicians who undertake processing and storage, growing out and maintenance in the field/glasshouse, multiplication and regeneration and characterisation/evaluation. Viability testing is undertaken at the seed and plant physiology laboratory at the Department of Plant Physiology. The curator keeps the 'central records' on the material (stock control and viability data) while research staff (including Ph.D students) are responsible for evaluation/ characterisation and crossing/selection experiments. Data on the latter are kept separately by research supervisors in the Department.

QUESTIONNAIRE: YOUR GENEBANK

NAME OF GENEBANK: GENEBANK OF ORNAMENTAL AND MEDICINAL PLANTS

LOCATION OF GENEBANK: DEPT. OF HORTIC., FACULTY OF AGRIC., X UNIVERSITY

YOUR NAME:

DATE: 1 / 12 / 92

SECTION I: YOUR GENEBANK'S RELATION WITH OTHER GENETIC RESOURCES PROGRAMMES

1. *Which category does your genebank fit into?*

- [x] Institutional
- [ ] National
- [ ] Regional
- [ ] International
- [ ] Other (please specify) _____

ANALYSIS

THE GENEBANK CONSERVES ONLY GERMPLASM THAT IS ACTUALLY OR POTENTIALLY USEFUL TO THE RESEARCH PROGRAMME AT THE UNIVERSITY.

2. *When was your genebank established?*

- [ ] In the planning stage
- [ ] Less than a year ago
- [ ] 1-2 years ago
- [ ] 3-5 years ago
- [x] More than 5 years ago

ANALYSIS

DIVERSE MANUAL DOCUMENTATION SYSTEMS (e.g. CATALOGUES) EXIST FOR SEVERAL PLANT SPECIES.

3. *Is there a national programme of plant genetic resources activities in your country?*

- [x] Yes or in the planning stage
- [ ] No

ANALYSIS

NATIONAL PROGRAM HAS BEEN INITIATED BUT DOCUMENTATION EXCHANGE BETWEEN OMP AND NP IS BEING PLANNED.

4. *Does your genebank collaborate with other genetic resource programmes in other institutes or genebanks?*

- [x] Yes
- [ ] No

ANALYSIS

COLLABORATION IS MAINLY THROUGH DEPARTMENTAL STAFF RESEARCH PROJECTS WITH OTHER RESEARCH INSTITUTIONS.

SECTION II: YOUR GENEBANK'S AIM AND AREAS OF ACTIVITIES

5. *Summarise in a short phrase your genebank's aim*

TO PROVIDE RESEARCHERS AND BREEDERS WITH STOCKS OF PRIMITIVE CULTIVARS, HYBRIDS, BREEDERS LINES AND MUTANTS, SO AS TO FURTHER RESEARCH AND SEED/SEEDLING PRODUCTION OF ORNAMENTAL AND MEDICINAL PLANTS.

[In this phrase try to explain what the genebank has been set up to achieve]

ANALYSIS

THE AIM IS CLOSELY LINKED TO THE UNIVERSITY'S RESEARCH INTERESTS.

6. *Which of the following activities does your genebank undertake? (tick as many as appropriate)*

- [x] Germplasm collecting
- [x] Acquisition of new germplasm samples
- [x] Multiplication/regeneration of germplasm
- [x] Characterisation and evaluation of germplasm
- [x] Maintenance of germplasm
- [ ] Germplasm health/quarantine
- [ ] Dissemination of information on germplasm — ON REQUEST. NOT ON A REGULAR BASIS
- [ ] Selection of germplasm for distribution
- [x] Research (e.g. germplasm enhancement, seed physiology)
- [ ] Organisation of technical meetings/training workshops

ANALYSIS

ACTIVITIES ARE UNDERTAKEN ACCORDING TO RESEARCH NEEDS OF DEPARTMENTAL STAFF AND THEIR POSTGRADUATE STUDENTS.

7. *In which form(s) is/are the germplasm maintained and in which type of collection?*

	Base	Active	Working
Seeds	☐	☒	☐
Field genebank/collection	☐	☒	☐
In vitro	☐	☐	☐
Pollen	☐	☐	☐

ANALYSIS

COLLECTIONS MAY BE FOUND IN THE FORM OF ORCHARD SEED (UNDER MEDIUM-TERM STORAGE CONDITIONS) AND FIELD.

8. *How many accessions are maintained by your genebank?*

- [ ] Less than 100
- [ ] 100-500
- [ ] 500-1,000
- [x] More than 1,000

ANALYSIS

ON AVERAGE, SAMPLES MAINTAINED UNDER MEDIUM-TERM STORAGE CONDITIONS CONTAIN 300 SEEDS/ACCESSION WITH FULL REGISTRATION DATA. THE NUMBER OF ACCESSIONS IN GENERAL WOULD DICTATE THE NECESSITY OF A COMPUTER FOR DATA MANAGEMENT.

9. *How many different species are held?*

- [ ] 1
- [ ] 2-10
- [x] More than 10

ANALYSIS

THE COLLECTION INCLUDES THE MAIN ORNAMENTAL AND MEDICINAL PLANTS (15-20 SPECIES) OF ECONOMIC IMPORTANCE. THE NUMBER OF SPECIES ADDS AN ADDITIONAL LEVEL OF DOCUMENTATION COMPLEXITY, SO A COMPUTER-

10. *Is your genebank in an expanding phase?*

Yes ☐
No ☒

IN TERMS OF STAFF AND STORAGE/GLASSHOUSE/FIELD FACILITIES, NO EXPANSION IS TAKING PLACE. DOCUMENTATION NEEDS ARE FORSEEN TO INCREASE, MAINLY IN MANAGEMENT STANDARDS, QUALITY & OUTP

11. *Approximately how many samples do you distribute in a typical year?*

Fewer than 100 ☒
100-500 ☐
More than 500 ☐

DISTRIBUTION DOES NOT MAKE A MAJOR DEMAND ON THE DOCUMENTATION SYSTEM.

SECTION III: YOUR GENEBANK'S ORGANISATION

12. *How does your genebank fit into the structure of your parent institute?*

Genebank is independent of other projects but sometimes works with other projects ☐
Genebank is independent of other projects but works closely with other projects ☐
Genebank is part of another project ☒
Other (please specify) ☐

GENEBANK IS PART OF THE ORNAMENTAL AND MEDICINAL SEED/SEEDLING PRODUCTION PROJECT.

GENEBANK IS PART OF THE DEPARTMENTAL RESEARCH PROJECT TO IMPROVE SEED/SEEDLING PRODUCTION OF ORNAMENTAL AND MEDICINAL PLANTS. DOCUMENTATION IS THEREFORE GUIDED BY RESEARCHERS NEEDS AND THE NEEDS OF THEIR COLLABORATOR DOCUMENTATION INPUT IS NOT RELATED TO OUTSIDE ORGANIZATIONS.

13. *Does your genebank work on a collaborative basis with other research stations (e.g. for the purposes of regeneration or evaluation of accessions?)* ***A*** *N A L Y S I S*

Occasionally ☐
Frequently ☐
No. never ☒

THE UNIVERSITY DOES NOT HAVE OTHER RESEARCH STATIONS.

If the answer to question 13 is "no, never", go straight to question 17, otherwise complete questions 14 & 15.

14. *How is the work organised at the different research stations?* ***A*** *N A L Y S I S*

By your genebank ☐
In consultation with your genebank ☐
Independently of your genebank ☐

15. *Do the different research stations operate, or plan to operate, documentation systems?*

A N A L Y S I S

Yes ☐
No ☐

If the answer to question 15 is "no", go straight to question 17, otherwise complete question 16.

16. *Have these documentation systems been developed independently of your genebank?*

A N A L Y S I S

Yes ☐
No ☐

17. *Which of the following statements most accurately describes the organisation of duties at your genebank?*

A N A L Y S I S

People do not have clearly defined duties and share duties on a day to day basis ☐
People have clearly defined duties but there is a regular sharing of duties on a day to day basis ☐
People have clearly defined duties but perform other duties when necessary ☒
People have clearly defined duties with little or no sharing of duties ☐

TECHNICAL STAFF UNDERTAKE ROUTINE GENEBANK DUTIES, IN ADDITION TO PROVIDING ASSISTANCE IN REGULAR EXPERIMENTAL WORK WHEN REQUIRED. THE DOCUMENT-ATION SYSTEM IS ENVISAGED TO BE USED BY MORE THAN ONE PERSON WHO MAY NOT BE WELL TRAINED.

18. *In which areas is there usually a sharing of duties?*
Indicate the approximate number of people involved

A N A L Y S I S

		Number:
Germplasm collecting	☐	
Acquisition of new germplasm samples	☒	3
Germplasm health/quarantine	☐	
Sample registration	☒	2
Seed cleaning	☒	5
Seed drying	☐	
Seed moisture content testing	☐	
Seed viability testing	☐	
Seed packing and storage	☒	5
Distribution of germplasm	☐	
Multiplication/regeneration of germplasm	☒	4
Dissemination of information on germplasm	☐	
Selection of germplasm for distribution	☐	
Characterisation and evaluation of germplasm	☒	4
Research (e.g. germplasm enhancement, seed physiology)	☒	6
Organisation of technical meetings/training workshops	☐	

THE LARGE NUMBER OF PEOPLE INVOLVED IN GENE-BANK TASKS PUTS RELATIVE EMPHASIS ON THE NEED FOR A FLEXIBLE, RELIABLE AND OPEN DOCUMENTATION SYSTEM. RESEARCHERS NEEDS ARE IMPORTANT.

19. *In which areas is there usually backlog of documentation work?*
Indication the extent (e.g. small, medium, large)

A N A L Y S I S

DATA IN FIELD RECORDS/ SHEETS FILLED FOR EXPERIMENTAL PLOTS/ DATA FOR EVALUATION AND LABORATORY ANALYSIS RECORDS NEED TO BE PROCESSED TO SUIT THE GENEBANK DOCUMENTATION CATALOGUES. INSUFFICIENT MANPOWER DELAYS THIS PROCESS. POTENTIALLY, A LARGE AMOUNT OF DATA WILL BE GENERATED. THE DOCUMENTATION SYSTEM NEEDS TO FACILITATE THE RECORDING OF THIS.

		Extent:
Germplasm collecting	☐	
Acquisition of new germplasm samples	☐	
Germplasm health/quarantine	☐	
Sample registration	☐	
Seed cleaning	☐	
Seed drying	☐	
Seed moisture content testing	☒	MEDIUM
Seed viability testing	☒	MEDIUM
Seed packing and storage	☐	
Distribution of germplasm	☐	
Multiplication/regeneration of germplasm	☒	LARGE
Dissemination of information on germplasm	☐	
Selection of germplasm for distribution	☐	
Characterisation and evaluation of germplasm	☒	MEDIUM
Research (e.g. germplasm enhancement, seed physiology)	☒	LARGE

20. *What are the genebank's documentation priorities?*
Indicate the degree (e.g. low, medium, high)

A N A L Y S I S

DEMAND IS HIGH FOR GOOD QUALITY INFORMATION TO BE HELD AND EXCHANGED. THE WAY IN WHICH DATA IS MANAGED MUST BE A PRIORITY IN ALL DOCUMENTATION WORK.

		Degree:
Germplasm collecting	☒	LOW
Acquisition of new germplasm samples	☒	HIGH
Germplasm health/quarantine	☐	
Sample registration	☒	HIGH
Seed cleaning	☐	
Seed drying	☐	
Seed moisture content testing	☐	
Seed viability testing	☐	
Seed packing and storage	☐	
Distribution of germplasm	☐	
Multiplication/regeneration of germplasm	☒	MEDIUM
Dissemination of information on germplasm	☒	HIGH
Selection of germplasm for distribution	☐	
Characterisation and evaluation of germplasm	☒	HIGH
Research (e.g. germplasm enhancement, seed physiology)	☒	MEDIUM
Organisation of technical meetings/training workshops	☐	

GENEBANK PROCEDURE ANALYSIS

NAME OF GENEBANK: GENEBANK OF ORNAMENTAL AND MEDICINAL PLANTS (OMP)

LOCATION OF GENEBANK: DEPT. OF HORTIC., FACULTY OF AGRIC., X UNIVERSITY

YOUR NAME:

DATE: 1/12/92

PROCEDURE: REGISTRATION

TYPE OF PROCEDURE (OPERATIONAL/SCIENTIFIC): SCIENTIFIC

RELATED PROCEDURES:

SHEET NUMBER: 1

DESCRIPTOR	DESCRIPTOR STATES	COMMENTS
ACCESSION NUMBER	GENEBANK NUMBER	CODING SYSTEM: OMPØØØØ — OMP9999
SPECIES NAME	SCIENTIFIC NAME OF SPECIES	FULL LATIN NAME
VARIETY NAME		VERNACULAR NAME
SOURCE	1, 2, 3 ..... etc.	1 = COLLECTING MISSION 2 = DONOR INSTITUTE, etc.
ORIGIN (COUNTRY/REGION)	e.g. 'AFG' = AFGHANISTAN	USE INTERNATIONAL STANDARD CODES FOR THE REPRESENTATION OF NAMES OF COUNTRIES (ISO)
ACQUISITION DATE	DAY/MONTH/YEAR	00/00/0000 e.g. 01/12/1992
DONOR NUMBER		AS RECEIVED

COMMENTS: This information is recorded in a manual documentation system in three separate files: (1) Registration; (2) Stock control; (3) Experimentation (Characterisation/Evaluation).

GENEBANK PROCEDURE ANALYSIS

CONTINUATION SHEET

PROCEDURE: CHARACTERISATION | EVALUATION — GENERAL

SHEET NUMBER: 2

DESCRIPTOR	DESCRIPTOR STATES	COMMENTS
ACCESSION NUMBER	GENEBANK NUMBER	CODING SYSTEM: OMPϕϕϕϕ – OMP9999
EXPERIMENTAL PLOT		NAME OF LOCATION
RESEARCHERS NAME		
SOWING | PLANTING DATE	DAY | MONTH | YEAR	e.g. 01|12|1992
CULTIVATION DESIGN		DISTANCES BETWEEN ROWS | LINES
AGRONOMIC CULTURAL PRACTICES		FERTILIZATION, IRRIGATION, WEEDING OUT, etc.
SOIL CHARACTERISTICS		pH, TYPE, etc.
DAYS TO FLOWERING		NUMBER OF DAYS FROM PLANTING TO 50% OF PLANTS FLOWERING
FLOWER COLOUR	0, 1, 2, 3, 4, 5, 6, 7, 8, 9	0- INTERMEDIATE REGION ABSENT; 1- WHITE; 2- YELLOW; 3- ORANGE; 4- PINK; 5- RED; 6- GREEN; 7- BLUE; 8- VIOLET; 9- BROWN
SEED HARVEST DATE	DAY | MONTH | YEAR	e.g. 01|12|1992
% OIL CONTENT (SEED)		PER CENTAGE | gm. wt.

COMMENTS:

Appendix II: Country codes

The International Standard codes (ISO) for the representation of names of countries

ENGLISH:	SPANISH:	FRENCH:	CODE:
A			
Afghanistan	Afganistán	Afghanistan	**AFG**
Albania	Albania	Albanie	**ALB**
Algeria	Argelia	Algérie	**DZA**
American Samoa	Samoa Americana	Samoa Américaines	**ASM**
Andorra	Andorra	Andorre	**AND**
Angola	Angola	Angola	**AGO**
Anguilla	Anguila	Anguilla	**AIA**
Antarctica	Antártida	Antarctique	**AT A**
Antigua and Barbuda	Antigua y Barbuda	Antigua-et-Barbuda	**ATG**
Argentina	Argentina	Argentine	**ARG**
Armenia	Armenia	Arménie	**ARM**
Aruba	Aruba, Isla	Aruba	**ABW**
Australia	Australia	Australie	**AUS**
Austria	Austria	Autriche	**AUT**
Azerbaijan	Azerbaiyán	Azerbaïdjan	**AZE**
B			
Bahamas	Bahamas	Bahamas	**BHS**
Bahrain	Bahrein	Bahreïn	**BHR**
Bangladesh	Banglasdesh	Bangladesh	**BGD**
Barbados	Barbados	Barbade	**BRB**
Belarus	Belarlús	Bélarus	**BLR**
Belgium	Bélgica	Belgique	**BEL**
Belize	Belice	Belize	**BLZ**
Benin	Benin	Bénin	**BEN**
Bermuda	Bermudas	Bermudes	**BMU**
Bhutan	Bhután	Bhoutan	**BTN**
Bolivia	Bolivia	Bolivie	**BOL**
Bosnia and Herzegovina	Bosnia y Herzegovina	Boznie-Herzégovine	**BIH**

The English and French country names and the country code columns in this table have been reproduced directly from: *International Standard Codes for the representation of countries,* 3rd ed., 1988. ISO, International Organization for Standardization. They are current as of 2/12/92. Most of the Spanish country names have been reproduced directly from: *Names of Countries,* fao terminology bulletin 20/rev. 8, 1986. Food and Agriculture Organization of the United Nations.

ENGLISH:	SPANISH:	FRENCH:	CODE:
Botswana	Botswana	Botswana	BWA
Bouvet Island	Bouvet, Isla	Bouvet, Ile	BVT
Brazil	Brasil	Brésil	BRA
British Indian Ocean Territory	Territoria Británico de1 Océano Indico	Océan Indien, Territoire britannique de 1'	IOT
British Virgin Islands	Vírgenes Británicas, Islas	Vierges (Britanniques), Iles	VGB
Brunei	Brunei Darussalam	Brunéi Darussalam	BRN
Bulgaria	Bulgaria	Bulgarie	BGR
Burkina Faso	Burkina Faso	Burkina Faso	BFA
Burundi	Burundi	Burundi	BDI
C			
Cambodia	Camboya	Cambodge	KHM
Cameroon	Camerún	Cameroun	CMR
Canada	Canadá	Canada	CAN
Cape Verde	Cabo Verde	Cap-Vert	CPV
Cayman Islands	Caimán, Islas	Caïmanes, Iles	CYM
Central African Republic	República Centroafricana	République centrafricaine	CAF
Chad	Chad	Tchad	TCD
Chile	Chile	Chili	CHL
China	China	Chine	CHN
Christmas Island (Australia)	Christmas, Isla	Christmas, Ile	CXR
Cocos (Keeling) Islands	Cocos (Keeling), Islas	Cocos (Keeling), Iles	CCK
Colombia	Colombia	Colombie	COL
Comoros	Comoras	Comores	COM
Congo	Congo	Congo	COG
Cook Islands	Cook, Islas	(look, Iles	COK
Costa Rica	Costa Rica	Costa Rica	CRI
Cote d'Ivoire	Côte d'Ivoire	Côte d'lvoire	CIV
Croatia	Croacia	Croatie	HRV
Cuba	Cuba	Cuba	CUB
Cyprus	Chipre	Chypre	CYP
Czechoslovakia	Checoslovaquia	Tchécoslovaquie	CSK
D			
Democratic People's Republic of Korea	Corea, República Popular Democrática de	Corée, République pop. démocratique de	PRK
Denmark	Dinamarca	Danemark	DNK
Djibouti	Djibouti	Djibouti	DJI
Dominica	Dominica	Dominique	DMA
Dominican Republic	República Dominicana	République dominicaine	DOM

ENGLISH:	SPANISH:	FRENCH:	CODE:
East Timor	Timor Oriental	Timor Oriental	**TMP**
Ecuador	Ecuador	Equateur	**ECU**
Egypt	Egipto	Egypte	**EGY**
El Salvador	El Salvador	El Salvador	**SLV**
Equatorial Guinea	Guinea Ecuatorial	Guinée équatoriale	**GNQ**
Estonia	Estonia	Estonie	**EST**
Ethiopia	Etiopía	Ethiopie	**ETH**
Falkland Islands (Malvinas)	Malvinas (Falkland), Islas	Falkland (Malvinas), Iles	**FLK**
Faroe Islands	Feroe, Islas	Féroé, Iles	**FRO**
Fiji	Fiji	Fidji	**FJI**
Finland	Finlandia	Finlande	**FIN**
France	Francia	France	**FRA**
French Guiana	Guayana Francesa	Guyane française	**GUF**
French Polynesia	Polinesia Francesa	Polynésie fraçaise	**PYF**
French Southern Territories	Tierras Australes Francesas	Terres australes françaises	**ATF**
Gabon	Gabón	Gabon	**GAB**
Gambia	Gambia	Gambie	**GMB**
Georgia	Georgia	Géorgie	**GEO**
Germany	Alemania	Allemagne	**DEU**
Ghana	Ghana	Ghana	**GHA**
Gibraltar	Gibraltar	Gibraltar	**GIB**
Greece	Grecia	Grèce	**GRC**
Greenland	Groenlandia	Greenland	**GRL**
Grenada	Granada	Grenade	**GRD**
Guadeloupe	Guadalupe	Guadeloupe	**GLP**
Guam	Guam	Guam	**GUM**
Guatemala	Guatemala	Guatemala	**GTM**
Guinea	Guinea	Guinée	**GIN**
Guinea-Bissau	Guinea-Bissau	Guinée-Bissau	**GNB**
Guyana	Guyana	Guyana	**GUY**
Haiti	Haiti	Haïti	**HTI**
Heard and Mc Donald Islands	Heard y Mc Donald, Islas	Heard et Mc Donald, Iles	**HMD**
Honduras	Honduras	Honduras	**HND**
Hong Kong	Hong Kong	Hong-Kong	**HKG**

ENGLISH:	SPANISH:	FRENCH:	CODE:
Hungary	Hungría	Hongrie	HUN
I			
Iceland	Islandia	Islande	ISL
India	India	Inde	IND
Indonesia	Indonesia	Indonésie	IDN
Iran	Irán	Iran	IRN
Iraq	Iraq	Iraq	IRQ
Ireland	Irlanda	Irlande	IRL
Israel	Israel	Israël	ISR
Italy	Italia	Italie	ITA
J			
Jamaica	Jamaica	Jamaïque	JAM
Japan	Japón	Japon	JPN
Jordan	Jordania	Jordanie	JOR
K			
Kazakhstan	Kazajstán	Kazakhstan	KAZ
Kenya	Kenya	Kenya	KEN
Kiribati	Kiribati	Kiribati	KIR
Korea, Republic of	Corea, República de	Corée, République de	KOR
Kuwait	Kuwait	Koweït	KWT
Kyrgyzstan	Kirguistán	Kirghizistan	KGZ
L			
Lao People's Democratic Republic	Lao, República Democrática Popular	Lao, République démocratique populaire	**LAO**
Latvia	Letonia	Lettónie	**LVA**
Lebanon	Líbano	Liban	**LBN**
Lesotho	Lesotho	Lesotho	**LSO**
Liberia	Liberia	Liberia	**LBR**
Libyan Arab Jamahiriya	Jamahiriya Arabe Libia	Libyenne, Yamahiriya Arabe	**LBY**
Liechtenstein	Liechtenstein	Liechtenstein	**LIE**
Lithuania	Lituania	Lituanie	**LTU**
Luxemburg	Luxemburgo	Luxembourg	**LUX**
M			
Macau	Macao	Macao	**MAC**
Madagascar	Madagascar	Madagascar	**MDG**
Malawi	Malawi	Malawi	**MWI**
Malaysia	Malasia	Malaisie	**MYS**
Maldives	Maldivas	Maldives	**MDV**
Mali	Malí	Mali	**MLI**

ENGLISH:	SPANISH:	FRENCH:	CODE:
Malta	Malta	Malte	MLT
Marshall Islands	Marshall, Islas	Marshall, Iles	MHL
Martinique	Martinica	Martinique	MTQ
Mauritania	Mauritania	Mauritanie	MRT
Mauritius	Mauricio	Maurice	MUS
Mexico	México	Mexique	MEX
Micronesia	Micronesia	Micronésie	FSM
Moldova, Republic of	Moldova, República de	Moldova, République de	MDA
Monaco	Monaco	Monaco	MCO
Mongolia	Mongolia	Mongolie	MNG
Montserrat	Monserrat	Monserrat	MSR
Morocco	Marruecos	Maroc	MAR
Mozambique	Mozambique	Mozambique	MOZ
Myanmar	Myanmar	Myanmar	MYA
N			
Namibia	Namibia	Namibie	NAM
Nauru	Nauru	Nauru	NRU
Nepal	Nepal	Nepal	NPL
Netherlands	Países Bajos	Pays-Bas	NLD
Netherlands Antilles	Antillas Neerlandesas	Antilles Néerlandaises	ANT
Neutral Zone	Zona Neutral	Zone Neutre	NTZ
New Caledonia	Nueva Caledonia	Nouvelle-Calédonie	NCL
New Zealand	Nueva Zelandia	NouvelIe-Zélande	NZL
Nicaragua	Nicaragua	Nicaragua	NIC
Niger	Níger	Niger	NER
Nigeria	Nigeria	Nigeria	NGA
Niue	Niue, Isla	Nioué	NIU
Norfolk Island	Norfolk, Isla	Norfolk, Iles	NFK
Northern Mariana Islands	Mariana de1 Norte, Islas	Mariannes du Nord, Iles	MNP
Norway	Noruega	Norvége	NOR
O			
Oman	Omán	Oman	OMN
P			
Pakistan	Pakistan	Pakistan	PAK
Palau	Palau, Islas	Palau	PLW
Panama	Panama	Panama	PAN
Papua New Guinea	Papua Nueva Guinea	Papouasie-Nouvelle-Guinée	PNG
Paraguay	Paraguay	Paraguay	PRY

ENGLISH:	SPANISH:	FRENCH:	CODE:
Peru	Perú	Pérou	PER
Philippines	Filipinas	Philippines	PHL
Pitcairn Islands	Pitcairn, Islas	Pitcairn	PCN
Poland	Polonia	Pologne	POL
Portugal	Portugal	Portugal	PRT
Puerto Rico	Puerto Rico	Porto Rico	PRI
Qatar	Qatar	Qatar	QAT
Reunion	Reunión	Réunion	REU
Romania	Rumania	Roumanie	ROM
Russian Federation	Federación de Rusia	Fédération de Russie	RUS
Rwanda	Rwanda	Rwanda	RWA
Saint Helena	Santa Elena	Sainte-Hélène	SHN
Saint Kitts and Nevis	Saint Kitt y Nevis	Saint-Kitts et Nevis	KNA
Saint Lucia	Santa Lucía	Sainte-Lucie	LCA
Saint Pierre and Miquelon	San Pedro y Miguelón	Saint-Pierre et Miquelon	SPM
Saint Vincent and the Grenadines	San Vicente y las Granadinas	Saint-Vincent et Grenadines	VCT
Samoa	Samoa	Samoa	WSM
San Marino	San Marino	Saint-Marin	SMR
Sao Tome e Principe	Santo Tomé y Príncípe	Sao Tomé-et-Principe	STP
Saudi Arabia	Arabia Saudita	Arabie saoudite	SAU
Senegal	Senegal	Sénégal	SEN
Seychelles	Seychelles	Seychelles	SYC
Sierra Leone	Sierra Leona	Sierra Leone	SLE
Singapore	Singapur	Singapour	SGP
Slovenia	Eslovenia	Slovénie	SVN
Solomon Islands	Salomón, Islas	Salomon, Iles	SLB
Somalia	Somalia	Somalie	SOM
South Africa	Africa de1 Sur	Afrique du sud	ZAF
Spain	España	Espagne	ESP
Sri Lanka	Sri Lanka	Sri Lanka	LKA
Sudan	Sudán	Soudan	SDN
Surinam	Suriname	Suriname	SUR
Svalbard and Jan Mayen Islands	Svalbard y Jan Mayen, Islas	Svalbard et Jan Mayen, Ile	SJM
Swaziland	Swazilandia	Swaziland	SWZ
Sweden	Suecia	Suède	SWE
Switzerland	Suiza	Suisse	CHE

ENGLISH:	SPANISH:	FRENCH:	CODE:
Syrian Arab Republic	Siria, República Arabe	Syrienne, République arabe	SYR
T			
Taiwan, Province of China	Taiwán, Provincia de China	Taiwan, province de Chine	TWN
Tajikistan	Tayikistán	Tadjikistan	TJK
Tanzania	Tanzania	Tanzanie	TZA
Thailand	Tailandia	Thaïlande	THA
Togo	Togo	Togo	TGO
Tokelau	Tokelau	Tokélaou	TKL
Tonga	Tongo	Tonga	TON
Trinidad and Tobago	Trinidad y Tabago	Trinité-et-Tobago	TTO
Tunisia	Túnez	Tunisie	TUN
Turkey	Turquía	Turquie	TUR
Turkmenistan	Turkmenistán	Turkmenistan	TKM
Turks and Caicos Islands	Turcas y Caicos, Islas	Turques et Caïques, Iles	TCA
Tuvalu	Tuvalu	Tuvalu	TUV
U			
Uganda	Uganda	Ouganda	UGA
Ukraine	Ukrania	Ukraine	UKR
United Arab Emirates	Emiratos Arabes Unidos	Emirats arabes unis	ARE
United Kingdom	Reino Unido	Royaume-Uni	GBR
United States Misc. Pacific Islands	Islas varias de1 Pacífico (EE.UU.)	Diverses Iles du Pacifique (Etats-Unis)	UMI
United States of America	Estados Unidos de America	Etats-Unis d'Amérique	USA
Uruguay	Uruguay	Uruguay	URY
Uzbekistan	Uzbekistan	Uzbekistan	UZB
V			
Vanuatu	Vanuatu	Vanuatu	VUT
Vatican City State	Vaticano, Estado de la Ciudad de1	Vatican, Etat de la cité du	VAT
Venezuela	Venezuela	Venezuela	VEN
Viet Nam	Viet Nam	Viet Nam	VNM
Virgin Islands (US)	Vírgenes Islas (EE.UU.)	Vierges, des Etats-Unis, Iles	VIR
W			
Wallis and Fortuna Islands	Wallis y Futuna, Islas	Wallis et Futuna, Iles	WLS
Western Sahara	Sahara Occidental	Sahara Occidental	ESH
Y			
Yemen, Republic of	Yemen	Yemen	YEM
Z			
Zaire	Zaire	Zaïre	ZAR
Zambia	Zambia	Zambie	ZMB
Zimbabwe	Zimbabwe	Zimbabwe	ZWE

Index

A

B

C

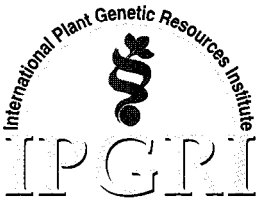
D

T

U

V

W



ISBN 92-9043-215-2